# TECHNIQUES OF STATISTICAL ANALYSIS

SELECTED

# TECHNIQUES OF

# STATISTICAL ANALYSIS

for Scientific and Industrial Research
and
Production and Management Engineering

BY THE

STATISTICAL RESEARCH GROUP, COLUMBIA UNIVERSITY

APPLIED MATHEMATICS PANEL, OFFICE OF SCIENTIFIC
RESEARCH AND DEVELOPMENT

EDITED BY

CHURCHILL EISENHART, National Bureau of Standards
MILLARD W. HASTAY, National Bureau of Economic Research
W. ALLEN WALLIS, University of Chicago

New York and London
McGRAW-HILL BOOK COMPANY, Inc.
1947

TECHNIQUES OF STATISTICAL ANALYSIS ᏏᏏ

IV

TO
WARREN WEAVER

# PREFACE

## NATURE OF THE BOOK

This book discusses a series of problems that occur frequently in planning, analyzing, or interpreting quantitative data. Various techniques appropriate to these problems are explained, in terms of both general principles and specific procedures.

The book is written for scientists, engineers, managers, and statisticians. It seeks to explain the statistical approach to various problems, to expound the theory that can be brought to bear on each problem, to point out the features of the problem that determine the relevance of the theory, to give detailed and workable computational procedures, including formulas, tables, and charts, and to illustrate the interpretation of the results. Concrete illustrative material is used to bring out the interplay of substantive and statistical considerations, the illustrations being selected primarily for the light they throw on the art of applying statistical techniques and only secondarily for intrinsic interest. Although relatively little new statistical theory is included, several topics are given more thoroughly integrated and carefully expounded treatment than has previously been available, and many new tables are included.

In preparing this book, the authors have kept in mind the reader who is interested in applying statistical techniques in his work but who either does not have much interest in or time for studying general statistical theory as such, or else feels that he can best understand and appreciate statistical theory through a series of special problems. Even the statistical theorist, however, should be able to deepen and broaden his understanding and appreciation of statistical theory by examining special problems in some detail.

The chemical, physical, and engineering sciences are, on the whole, distinctly backward in the statistical planning of experiments and the analysis of empirical data, at least by comparison with the social, biological, and agricultural sciences. No doubt this is partly because physical scientists have attained a high degree of proficiency in experimental techniques and can make repeated, independent experimental investigations of a given phenomenon—they can "think with their hands"—so that competent investigators are likely to learn the truth (though perhaps belatedly and at excessive cost) despite crudities in statistical techniques. In the social and biological sciences, on the other hand, repetition or control of experimental procedures is usually impossible, at least in anything like the degree achieved in the physical sciences, so that the truth can be learned only through highly developed statistical techniques closely coupled (as statistical techniques must always be) with sound scientific knowledge and techniques. But even though sound statistical techniques may not be indispensable in the physical sciences, they are invaluable. Increasing recognition of this, accelerated by developments during the war, is leading to the rapid introduction of statistical methods into chemistry, physics, and engineering; indeed, it seems likely that in the immediate future the greatest contributions of statistics will be to, and its greatest developments will arise from, the physical sciences. (The physical sciences may, however, yield this distinction to medicine, which is said to be nearly as dependent on statistics as the social sciences and nearly as backward statistically as the physical sciences.)

## BACKGROUND

During the war the Statistical Research Group carried on three kinds of work: (1) It advised and assisted the Army, the Navy, and the Office of Scientific Research and Development on the statistical aspects of problems arising in their activities; this constituted the largest part of the work. (2) It assumed primary responsibility for the investigation of certain problems of a predominantly statistical or probabilistic character. (3) It developed, reviewed, or expounded statistical techniques; this was the smallest part of the work and was incidental to the first two.

The great bulk of the work fell into the first two categories, and the final products did not emphasize statistical techniques as such. From this work, however, there arose a number of developments in applied statistics, as well as some in theoretical statistics. This book has been prepared to preserve and to disseminate useful statistical material that was almost a by-product of the responsible and cooperative handling of practical military, scientific, engineering, and production problems.

## AUTHORSHIP

Because it would be difficult or impossible accurately to assign personal credit and responsibility for this volume, its authorship is attributed to the Statistical Research Group as a whole. Under ordinary circumstances the basic material of the various chapters would have been published in technical journals by the originators individually. Under the actual circumstances, many ideas are presented here for the first time in chapters prepared by others than those who did the basic work. It thus becomes important to assign credit for the ideas in such chapters without at the same time implying responsibility for the entire content or emphasis. An attempt has been made to do this by notes at the beginning of the chapters. The editors, and in particular the undersigned, however, have exercised final judgment and authority over the form and content of each chapter and are thus in the unhappy position of having to assume all responsibility for the shortcomings of the book without being able to claim credit for its merits.

The following were members of the senior scientific staff of the Statistical Research Group, and the influence of each is reflected in the book.

Kenneth J. Arnold, *University of Wisconsin*
Rollin F. Bennett, *Booz, Allen and Hamilton*
Julian H. Bigelow, *Institute for Advanced Study*
Albert H. Bowker, *Stanford University*
Churchill Eisenhart, *National Bureau of Standards*
H. A. Freeman, *Massachusetts Institute of Technology*
Milton Friedman, *University of Chicago*
M. A. Girshick, *Stanford University*
Millard W. Hastay, *National Bureau of Economic Research*
Harold Hotelling, *University of North Carolina*
Edward Paulson, *University of Washington*
L. J. Savage, *University of Chicago*
Herbert Solomon, *Office of Naval Research*
George J. Stigler, *Columbia University*
A. Wald, *Columbia University*
W. Allen Wallis, *University of Chicago*
J. Wolfowitz, *Columbia University*

Of the junior scientific staff, those whose work has especially affected this volume include Nancy Brixey Breen, Henry Goldberg, Anne Frankenthaler Kohn, Harriet Levine, and Myra Levine.

## ACKNOWLEDGMENTS

A large share of any credit for the work of the Statistical Research Group must be assigned to Warren Weaver, Chief of the Applied Mathematics Panel, for his rare talents in administering research; he succeeded on the one hand in providing the information, guidance, and evaluation necessary to steer the work into channels effectively furthering the objectives of the Office of Scientific Research and Development and, on the other hand, in allowing the wide range of individual freedom, initiative, and responsibility that is vital to research.

Preparation and publication of this volume would probably have been impossible, and certainly would have been most difficult, without the indefatigable, efficient, and unstinting assistance of Mina Rees of the Applied Mathematics Panel. George B. Pegram, Chairman of the Columbia University Committee on War Research, gave encouragement and material assistance that was important in launching and carrying through the project to prepare this volume.

On nearly every chapter Kenneth J. Arnold gave the editors advice and assistance, which eliminated a number of errors and in many cases introduced substantial improvements.

Special efforts have been made to insure the accuracy of reference tables   Through the courtesy of the National Bureau of Standards, proofs of Tables 2.1, 8.1, 8.2, 8.3, 8.4, 15.1, and 15.2 were checked by the New York Mathematical Tables Project directed by Arnold N. Lowan; the printed tables were tested by differencing or, when necessary, by recomputation or other means. Proofs of Tables 1.1, 1.5, 14.1, and 16.4 were compared with published tables not used in their preparation.

The index was prepared by Churchill Eisenhart, assisted by Celia S. Martin.   Edythe B. Rodermund prepared and, insofar as time permitted, edited the typescript with extraordinary diligence, accuracy, and thoroughness.   The charts (except Figs. 10.1 and 10.2) are the work of H. Irving Forman.

W. ALLEN WALLIS
*Director of Research*
*Statistical Research Group*

CHICAGO, ILL.,
*January,* 1947.

# CONTENTS*

* An analytical table of contents appears at the beginning of each chapter.

xi

# LIST OF REFERENCE TABLES

# LIST OF REFERENCE CHARTS

## *Part* I

## INDUSTRIAL STATISTICS

## CHAPTER 1

## USE OF VARIABLES IN ACCEPTANCE INSPECTION FOR PERCENT DEFECTIVE

TABLES

This chapter was prepared by
W. Allen Wallis, University of Chicago

Secs. 11.4, 11.5, and 12 were drafted by Harriet Levine.   The material of
Secs. 10.2.3 and 11.1.4 is due to Kenneth J. Arnold.   The comparison made
in Sec. 5.2 was suggested by Milton Friedman.   The general character of
the Statistical Research Group's work on inspection by variables was
greatly influenced by M. A. Girshick, who had direct charge of the work
in which this type of inspection was most used.

# CHAPTER 1

## USE OF VARIABLES IN ACCEPTANCE INSPECTION FOR PERCENT DEFECTIVE

### 1. ATTRIBUTES AND VARIABLES

Acceptance inspection is the process of judging whether a lot of items is of acceptable quality by examining a sample from the lot. Such inspection is widely and successfully used in industry. It is usually carried out by classifying each item in the sample as either defective or nondefective, then deciding from the number of defective items in the sample whether the percentage of defective items in the lot is small enough for the lot to be considered acceptable. Thus, the quality of the lot is measured by the percentage of defective items, and this is judged by classifying the sample items according to attributes.

Often the sample items are classified as defective or nondefective on the basis of a measurable quality characteristic. For example, the items may be classified as too heavy or not, although the actual weight could be measured; they may be classified as too short or not, although the length could be measured; they may be classified as having an excessive volatile content or not, although the volatile content could be measured; they may be classified as operating too fast or not, although the speed could be measured; they may be classified as having too short a useful life or not, although the useful life could be measured. Such cases, where inspection is based on attributes, may be contrasted with cases in which inspection is based on variables, that is, on the actual measured values of the characteristic.

Inspection by variables has one major advantage over inspection by attributes: fewer items need be inspected for a given degree of accuracy in judging a lot's acceptability; or, conversely, decisions about the lot are more reliable if based on a sample inspected by variables than if based on a sample of the same size inspected by attributes. The reason for this is that variables inspection provides more information about the quality of each item; for example, that the volatile content is 16.02 percent or 20.61 percent, as the case may be, instead of merely that the volatile content exceeds, say, 16 percent in both cases. Since variables inspection provides more information about each item, a smaller number of sample items gives the same amount of information about the lot, or the same number of sample items gives more information about the lot.

The principal advantage of inspection by attributes over inspection by variables is that it requires less skill, less time, or less expensive equipment to inspect each item; less record-keeping to record the results of inspection; less arithmetic to determine from the sample results whether the lot should be judged acceptable. Usually this advantage of simplicity more than offsets the fact that more items must be inspected under attribute than under variables inspection.

Another advantage of attribute inspection is that the selection and installation of sound sampling plans is easier. In fact, complete and detailed procedures are available in such form that little or no previous knowledge of technical statistics is required for their successful use.* In attribute inspection the mathematical theory about the manner in which quality varies from one item to another is always fulfilled if the sample is selected at random from the lot. The mathematical theory underlying variables inspection plans, on the other hand, is that quality varies among the items of the lot in accordance with a normal (Gaussian) distribution; whether this theory is a sufficiently good approximation to the facts in a particular inspection problem is a matter requiring investigation by statistical methods.

There are other less important reasons for preferring attributes to variables. One of these is a misapprehension as to the amount of arithmetic required, often based in large part on ignorance of short-cut computational procedures. Another such reason is exaggeration of the limitations imposed by the normality assumption in the mathematics underlying variables inspection.

Attributes are sometimes used instead of variables because well-known sampling plans for variables measure lot quality by the average or the variability of the measurements, while well-known attribute plans measure it by the percent defective. Percent defective is often the appropriate measure, and its appropriateness is not affected by shifting between attributes and variables. For example, in inspecting rods for length, the percent too short may in some circumstances be a better measure of quality than either the mean or the variability of length—as when rods below a certain length are useless while those above that length can easily be trimmed at virtually no extra cost. Similarly, in determining whether individual packages in a lot provide full measure, the percent below the specified measure may be more significant than the average measure. Variables plans for percent defective exist but are not widely known.†

* See, for example, Statistical Research Group, *Sampling Inspection;* Dodge and Romig, *Sampling Inspection Tables* (for full citations see Sec. 13).

† See especially the following two papers, of which this chapter is primarily a development: JENNETT and WELCH, "The Control of Proportion Defective"; JOHNSON and WELCH, "Applications of the Non-central *t* Distribution."

Although in most cases attribute plans are undoubtedly more economical than variables plans for percent defective, because the lower cost of inspecting each item usually more than offsets the increase in the number of items that must be inspected, variables inspection offers decided advantages in far more instances than has usually been recognized. This is especially true when the inspection process is costly, as for example when it requires expensive time, equipment, or personnel, or when the items are expensive and are damaged or destroyed by inspection. It is also true when measurement by variables costs only a little more than classification by attributes, or when for purposes other than acceptance inspection the data on variables are more valuable than those on attributes. Thus it is desirable that the practical aspects of variables plans be more fully developed and the plans made more generally available than has heretofore been the case.

## 2. BACKGROUND OF THE PLAN

A variables plan of acceptance inspection for percent defective is based upon the sample mean and the sample standard deviation. Other measures of central tendency and variability could be used, for example, the median and the range. No other measures are considered in this chapter, however, because other measures all require larger samples to give as good results as the mean and standard deviation. The gains in computational simplicity that may be afforded by other measures are likely to be unimportant in the situations for which variables inspection is appropriate. Although in routine inspection operations it is not necessary to compute either the mean or the standard deviation, it will clarify the ideas to begin the discussion in terms of them.

A normal distribution, such as that shown in Fig. 1.1, is characterized by two quantities (parameters), its mean $\mu$ and its standard deviation $\sigma$. Normal distributions have large or small means according as $\mu$ is large or small; the observations are tightly clustered or widely scattered about the mean according as $\sigma$ is small or large. The proportion of the observations exceeding a certain upper limit, say $U$, depends upon how far the limit is above the mean, measured in terms of the standard deviation. For example, if the limit exceeds the mean by twice the standard deviation, that is, if $U = \mu + 2\sigma$, 2.275 percent of the observations exceed the limit; but if the limit exceeds the mean by three times the standard deviation, that is, if $U = \mu + 3\sigma$, only 0.135 percent of the observations exceed the limit. To determine the percent of observations exceeding any limit, first subtract the mean from the limit and then divide the difference by the standard deviation; that is, compute

$$K = \frac{U - \mu}{\sigma} \qquad (1)$$

which shows the difference between the limit and the mean as a multiple of the standard deviation. Various tables show for different values of $K$ what part of a normal distribution exceeds $K$.*
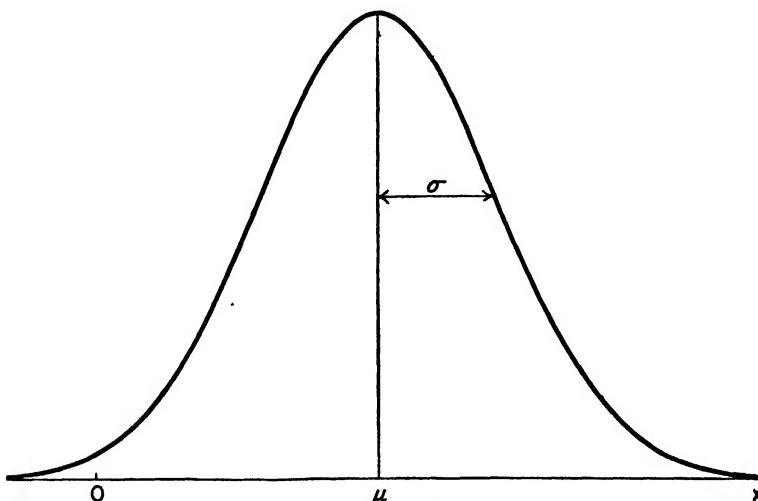


FIG. 1.1.—Normal distribution.

From formula (1) it is clear that the proportion exceeding a given limit $U$, which we will call the fraction defective, depends upon both the mean and the standard deviation. For a given standard deviation, the fraction defective is smaller the smaller the mean; but, however small the mean, the standard deviation may be large enough to cause a high

* See, for example, Table 1.1 in this chapter. By far the most extensive table is that prepared by the New York Mathematical Tables Project, *Tables of Probability Functions.* This gives probabilities to 15 decimal places for values of $K$ (denoted by $x$ in the tables) from 0 to 1 by steps of 0.0001, and for all values of $K$ above 1 by steps of 0.001.

Some care is required in using tables of the normal probability function, since two different kinds of table are common and several methods of tabulation are used. The two kinds of table differ in the units used to measure the difference between the limit and the mean, the most common type (illustrated by Vol. II of the NYMTP tables) measuring this as a multiple of the standard deviation and the other type (illustrated by Vol. I of the NYMTP tables) measuring it as a multiple of $\sqrt{2}$ times the standard deviation. Thus, to enter the second type of table, $K$ computed from formula (1) must be divided by $\sqrt{2} = 1.414$. In either type of table the entries may represent (a) the proportion exceeding $K$, which is the same as the proportion below $-K$; (b) the proportion below $K$, which is the same as the proportion above $-K$; (c) the proportion between the mean and $K$, which is the same as the proportion between the mean and $-K$; (d) the proportion between $-K$ and $K$; (e) the proportion below $-K$ and above $+K$. Given any one of these five values, the others are easily computed from the facts that the distribution is symmetrical about the mean and that the total proportion is one.

fraction defective. For a fixed mean less than the limit, the fraction defective is smaller the smaller the standard deviation; but, however small the standard deviation, the mean may be large enough to cause a high fraction defective. The acceptability of any mean, when quality is measured by fraction defective, depends on the standard deviation; and the acceptability of any standard deviation depends on the mean.

Figure 1.2 illustrates various combinations of the mean and standard deviation that correspond to the same lot quality, as measured by fraction
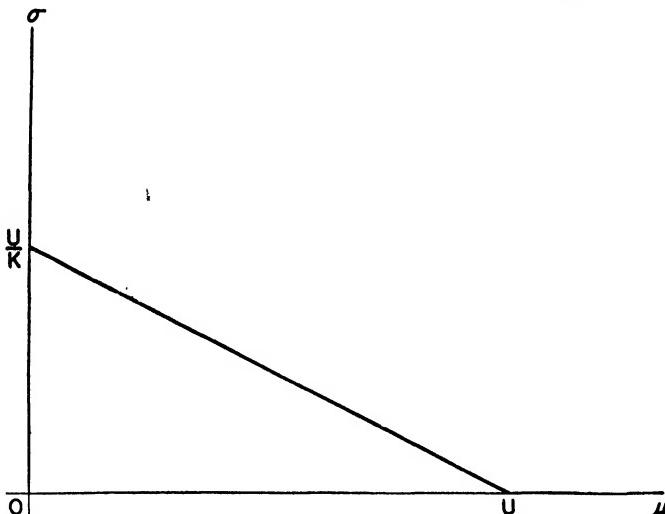


FIG. 1.2.—Combinations of $\mu$ and $\sigma$ corresponding with the same percent defective.

defective. Any combination corresponding to a point on the sloping line represents the same quality; combinations to the left of the line represent better quality (lower fraction defective), those to the right worse quality. Any lot quality can be represented by such a line, the percent defective being greater the steeper the line.* Thus a lot is acceptable, in the sense that it has a fraction defective less than or equal to a specified acceptable quality level (AQL), whenever $(U - \mu)/\sigma$ is greater than or equal to a certain value of $K$ which depends on the AQL and can be determined from tables of the normal probability function. In other words, the lot is acceptable if its mean plus a certain multiple of its standard deviation does not exceed the limit; that is, if

$$\mu + K\sigma \leq U \qquad (2)$$

* Lot qualities worse than 50 percent defective would be represented by a line starting at $\mu = U$, $\sigma = 0$ and sloping upward to the right. When the mean exceeds the limit the percent defective exceeds 50, and for a given mean the percent defective is smaller the larger the standard deviation. Such cases are rarely of interest in acceptance sampling but can be treated by formulas (1'), (2'), and (3') below.

The discussion so far has been entirely in terms of the lot mean and standard deviation. In practice, the lot mean and standard deviation, $\mu$ and $\sigma$, are unknown, and it is necessary to judge from the sample mean and standard deviation, $\bar{x}$ and $s$, whether they satisfy criterion (2). This can be done by adding a multiple of the sample standard deviation to the sample mean and comparing the sum with the limit. In choosing the multiple of the sample standard deviation, however, allowance must be made for the fact that the mean and standard deviation determined from the sample are subject to fluctuations from sample to sample. That is, if a number of samples are drawn from lots having exactly the same fraction defective, the samples will not all yield the same value of $(U - \bar{x})/s$. Thus the acceptable level of $(U - \bar{x})/s$, say $k$, must be small enough so that there will not be too high a probability that an acceptable lot will fail to meet the acceptance criterion for the sample, namely,

$$\bar{x} + ks \leq U \tag{3}$$

If the sample contains measurements on $N$ items (that is, contains $N$ observations) represented by $x_1, x_2, \cdots x_N$, the sample mean $\bar{x}$ is

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_N}{N} = \frac{\Sigma x_i}{N} \tag{4}$$

the sample variance $s^2$ is

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_N - \bar{x})^2}{N - 1} = \frac{\Sigma(x_i - \bar{x})^2}{N - 1} \tag{5}$$

and the sample standard deviation $s$ is the square root of the sample variance.

A few miscellaneous remarks will complete the background of variables inspection plans for percent defective:

The symbols $\mu$ and $\sigma$ are used to indicate the population mean and standard deviation; $\bar{x}$ and $s$ are used to represent the sample mean and standard deviation. Similarly, $K$ is used to represent the deviation of the lot mean from the limit in units of the lot standard deviation, and $k$ is used to represent the deviation of the sample mean from the limit in units of the sample standard deviation. This is in conformity with the sound and increasingly prevalent practice of using notation that emphasizes the distinction between true *lot* (*population* or *universe*) *parameters* and *sample estimates* of these parameters—an estimate meaning in statistical usage a value calculated from a sample (not guessed or judged in the everyday sense of "estimate") according to established principles for estimating lot parameters from sample data. Neglect of the distinction between lot parameters and sample estimates of them can be (and often is) a cause of serious confusion. The quantities $\bar{x}$ and

$s$, which for brevity have been called the sample mean and standard deviation, should properly be called the sample estimates of the lot mean and standard deviation.

The use of $N - 1$ instead of $N$ in the denominator of the variance [see formula (5)] has no real influence on the sampling plan provided the $k$ of expression (3) is chosen with due regard to whether $N$ or $N - 1$ is used. If $N$ is used, the proper value of $k$ is greater by a factor $\sqrt{N/(N - 1)}$ than if $N - 1$ is used, so $k$ times $s$ is the same in either case. If, however, the value of $s$ is itself of interest, the use of $N - 1$ is preferable.

Although formula (5) is useful for bringing out the meaning of the standard deviation, it should not be used for computing purposes. For computing, use one of the following two forms:

$$s^2 = \frac{\Sigma x^2 - \dfrac{(\Sigma x)^2}{N}}{N - 1} \tag{6a}$$

$$= \frac{N \Sigma x^2 - (\Sigma x)^2}{N(N - 1)} \tag{6b}$$

For purposes of acceptance inspection, however, it is not necessary (as will be seen later) actually to compute the standard deviation.

For concreteness it is assumed that defective items are those for which the value of the quality characteristic exceeds some upper bound of satisfactory quality $U$. It would be equally appropriate to assume that defective items are those below some minimum $M$. In that case, equations (1), (2), and (3) would be

$$K = \frac{\mu - M}{\sigma} \tag{1'}$$

$$\mu - K\sigma \geq M \tag{2'}$$

$$\bar{x} - ks \geq M \tag{3'}$$

In what follows, the assumption is retained that defectives are items that are too large, but from equations (1'), (2'), and (3') the application to cases where defectives are items that are too small will be obvious.

## 3. THREE PHASES OF A VARIABLES SAMPLING PLAN

### 3.1. The Three Phases

A sampling procedure may be analyzed into three phases:

(1) The *plan of action*, that is, the set of rules on the basis of which to accept or reject the lot.*

(2) The *amount of inspection* required by the plan, that is, the number of items that must be inspected from each lot.

---

* Some plans allow a third decision, to inspect more items: see Sec. 11.6.

(3) The *operating characteristics* of the plan, that is, the proportions of submitted lots of various qualities that will be accepted and rejected if the plan is used.

## 3.2. Plan of Action

The plan of action for variables inspection for percent defective is as follows:

a. A sample of $N$ items is selected at random from the lot.

b. The quality characteristic being inspected is measured and recorded for each item of the sample.

c. The mean and standard deviation are computed from the $N$ measurements.

d. The sample standard deviation is multiplied by a certain constant $k$ and added to the sample mean.

e. The lot is rejected if the value of $\bar{x} + ks$ found in step $d$ exceeds $U$, the limit of the quality characteristic above which an item is considered defective; otherwise, the lot is accepted.

In practice, steps $c$ and $d$ can be replaced by a single step that consists essentially of adding the squared values of the $N$ sample measurements; in that case step $e$ remains essentially as stated, though the actual numbers used in it are different.

Thus, the plan of action is characterized by two parameters: the sample size $N$ and the multiplier of the standard deviation $k$. ($U$ is associated with the definition of quality rather than with the sampling plan, since it simply defines a defective item.)

## 3.3. Amount of Inspection

Under the plans considered here, which are single-sampling plans, the number of items inspected per lot is always the same for a given plan.

Single-sampling plans for fraction defective based on attributes can sometimes be curtailed, that is, terminated without inspecting all $N$ items. Curtailing is possible as soon as the number of defectives reaches the rejection number or the number of defectives plus the number of items remaining to be inspected is less than the rejection number. Plans based on variables can also be curtailed, provided there is some minimum possible value of the characteristic being inspected. Most acceptance inspection involves quantities that can under no circumstances be less than zero, so curtailing before inspecting all $N$ items is technically possible. The simplest example of this possibility is the case in which, with $k$ positive, the sum of fewer than $N$ measurements exceeds $NU$, making it clear that for the entire $N$ items $\bar{x}$ will exceed $U$ and $\bar{x} + ks$ will exceed it still further. That there are other cases, less extreme but more complicated to describe, in which curtailing is possible can be seen by studying Sec. 11.3 below. Ordinarily, however, curtailing is not considered

because computations are not started until all measurements have been made. When it is possible to carry out simple computations while the data are being collected and it is important to minimize the amount of inspection, sequential analysis of the data will be considerably more efficient than merely curtailing a single-sampling plan. A brief outline of a plan for sequential analysis of fraction defective by variables is presented in Sec. 11.6.

### 3.4. Operating Characteristics (OC)

There are evidently many possible plans of action, for each different combination of sample size and multiplier of the standard deviation
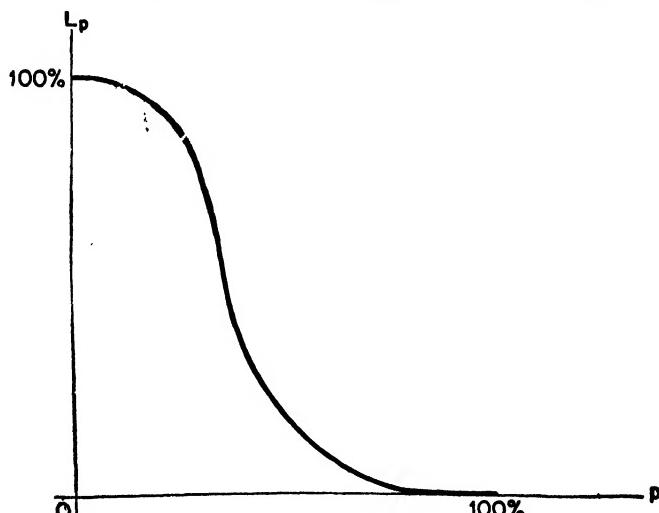


Fig. 1.3.—Typical operating-characteristic (OC) curve.

gives a different plan of action. If a large number of lots containing a given fraction defective $p$ are submitted to a given plan, a certain proportion $L_p$ are accepted; and this proportion differs from one plan to another. For each plan of action, it is important to know what proportion of submitted lots will be accepted for each possible quality $p$ of submitted lots, as a plan is clearly unsuitable if it passes too many of the lots of unsatisfactory quality or rejects too many of the lots of acceptable quality that are submitted to it.

This information is shown by the operating-characteristic (OC) curve of the plan. The ordinate of the OC curve shows the proportion $L_p$ of lots that will be accepted when the fraction defective in submitted lots is $p$. The proportion of submitted lots of quality $p$ that will be rejected is $1 - L_p$. The OC curve of Fig. 1.3 is typical: it shows a probability of 1 of accepting a lot having no defectives. As the fraction defective

increases, the probability of acceptance declines, slowly at first, then more rapidly, and finally flattens out.   It shows a probability of 0 of accepting a lot having only defectives.

It should be emphasized that the OC curve does *not* show the probability that an accepted lot will be of quality $p$; this depends upon the probability that a lot of quality $p$ will be submitted, as well as upon the probability that a lot of quality $p$ will be accepted *if* submitted.   The OC curve shows only the probability that a lot of quality $p$ will be accepted *if* submitted.

The OC curve may be thought of in another way, which, though it requires visualizing a three-dimensional chart, is illuminating.   Consider Fig. 1.2, in which the lot mean $\mu$ is plotted on the horizontal axis and the lot standard deviation $\sigma$ on the vertical axis.   Any point in this figure for which $\sigma \geq 0$ represents a possible lot.   Imagine a line perpendicular to the plane of the page through some point $\mu,\sigma$.   The height of the perpendicular represents the probability $L_{\mu,\sigma}$ that, if a lot whose mean and standard deviation correspond to the point $\mu,\sigma$ is submitted, the plan of action will accept it.   If an OC surface is constructed by erecting such perpendiculars at all possible combinations of $\mu$ and $\sigma$, it will be found that above any straight line in the $\mu$-$\sigma$ plane passing through $U$ (as for example, the sloping line in Fig. 1.2) all the values of $L_{\mu,\sigma}$ are equal. Since all lots represented by a straight line through $U$ have the same value of $K$, this means that all lots having the same value of $K$ have the same probability of acceptance.   Hence the OC surface can be reduced to an OC curve in two dimensions, the probability of acceptance being plotted against the value of $K$.   But, as was pointed out in Sec. 2, all lots having the same value of $K$ have the same fraction defective.   Thus the OC curve can show the probability of acceptance as a function of $p$, as in Fig. 1.3.

### 4. PLAN OF ACTION

#### 4.1. General Procedure

The first step in selecting the values of $k$ and $N$ that determine a particular plan is to select two points through which it is desired that the OC curve pass.   The first point corresponds to a lot fraction defective $p_1$ for which it is desired that the probability of rejection be not greater than $\alpha$.   The second point corresponds to a lot fraction defective $p_2$ $(p_2 > p_1)$ for which it is desired that the probability of acceptance be only $\beta$ $(\alpha + \beta < 1)$.   Fractions defective of $p_1$ or less thus represent acceptable quality; $\alpha$ is the maximum producer's risk, that is, the maximum risk that an acceptable lot will be rejected by the sampling plan. Similarly, fractions defective of $p_2$ or more represent unacceptable quality; $\beta$ is the maximum consumer's risk, that is, the maximum risk that

an unacceptable lot will be accepted by the sampling plan.   For lots of quality better than $p_1$ (fractions defective less than $p_1$) the producer's risk will be less than $\alpha$, and for lots of quality worse than $p_2$ (fractions defective more than $p_2$) the consumer's risk will be less than $\beta$.

The two points on the OC curve thus correspond with

$$p = p_1 \qquad L_{p_1} = 1 - \alpha \tag{7}$$
$$p = p_2 \qquad L_{p_2} = \beta \tag{8}$$

From these, $k$ and $N$ can be computed.   If $N$ proves too large, it is necessary to select two points on a less discriminating OC curve.   The more discriminating, that is, the steeper, the OC curve, the larger $N$; consequently, $N$ may be reduced by widening the gap between $p_1$ and $p_2$ or by increasing the sum of the risks $\alpha + \beta$, either of which results in a flatter OC curve

### 4.2. Formulas for $k$ and $N$

Let $K_\epsilon$, where $0 \leq \epsilon \leq 1$, represent a normal deviate exceeded with probability $\epsilon$; that is, $K_\epsilon$ is defined by

$$\frac{1}{\sqrt{2\pi}} \int_{K_\epsilon}^{\infty} e^{-\frac{x^2}{2}} dx = \epsilon \tag{9}$$

Values of $K_\epsilon$ corresponding with various common values of $\epsilon$ are given in Table 1.1.   Four values of $\epsilon$, namely, $p_1$, $p_2$, $\alpha$, and $\beta$, are of particular interest; and the corresponding values of $K_\epsilon$ may be designated $K_1$, $K_2$, $K_\alpha$, and $K_\beta$.

The value of $k$ may be computed from the formula

$$k = \frac{K_\alpha K_2 + K_\beta K_1}{K_\alpha + K_\beta} \tag{10}$$

The sample size can be determined either from

$$N = \frac{2(K_\alpha + K_\beta)^2 + (K_\alpha K_2 + K_\beta K_1)^2}{2(K_1 - K_2)^2} \tag{11}$$

or from

$$N = \frac{k^2 + 2}{2} \left( \frac{K_\alpha + K_\beta}{K_1 - K_2} \right)^2 \tag{12}$$

The value of $N$ given by (11) and (12) ordinarily will not be an integer; it should be rounded to the next higher integer, not to the nearest integer.*

Note that $k$ is a weighted average of $K_1$ and $K_2$, the weights being $K_\beta$ and $K_\alpha$, respectively.   Thus, if both $\alpha$ and $\beta$ are less than 0.5, $k$ is between $K_1$ and $K_2$; and if $\alpha$ and $\beta$ are equal, $k = (K_1 + K_2)/2$.

* Two more formulas for $N$ are given by equations (134) and (135), Sec. 11.1.2.

TABLE 1.1
NORMAL DEVIATES EXCEEDED WITH GIVEN PROBABILITY

Values of $K_\epsilon$ for various $\epsilon$, where $\dfrac{1}{\sqrt{2\pi}}\displaystyle\int_{K_\epsilon}^{\infty} e^{-\frac{x^2}{2}}\,dx = \epsilon$
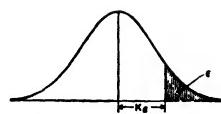
| $\epsilon$ | 0.00 | 0.0 | 0.1 | 0.2 | 0.3 | 0.4 |
|---|---|---|---|---|---|---|
| 0 | $\infty$ | $\infty$ | 1.28155 | 0.84162 | 0.52440 | 0.25335 |
| 1 | 3.09023 | 2.32635 | 1.22653 | 0.80642 | 0.49585 | 0.22754 |
| 2 | 2.87816 | 2.05375 | 1.17499 | 0.77219 | 0.46770 | 0.20189 |
| 3 | 2.74778 | 1.88079 | 1.12639 | 0.73885 | 0.43991 | 0.17637 |
| 4 | 2.65207 | 1.75069 | 1.08032 | 0.70630 | 0.41246 | 0.15097 |
| 5 | 2.57583 | 1.64485 | 1.03643 | 0.67449 | 0.38532 | 0.12566 |
| 6 | 2.51214 | 1.55477 | 0.99446 | 0.64335 | 0.35846 | 0.10043 |
| 7 | 2.45726 | 1.47579 | 0.95417 | 0.61281 | 0 33185 | 0.07527 |
| 8 | 2.40892 | 1.40507 | 0.91537 | 0.58284 | 0.30548 | 0.05015 |
| 9 | 2.36562 | 1.34076 | 0.87790 | 0.55338 | 0.27932 | 0 02507 |

When $\epsilon > 0.5$, $K_\epsilon = -K_{1-\epsilon}$.   When $\epsilon = 0.5$, $K_{0.5} = 0$.

The entries in column ($\epsilon$) are additional digits for the values of $\epsilon$ shown in the top row.   Thus, the entry 1.12639 in column (0.1) and row (3) corresponds with $\epsilon = 0.13$; similarly, the entry 2.74778 in column (0.00) and row (3) corresponds with $\epsilon = 0.003$.

A final $5$ in italic indicates that, in rounding, the nearest *odd* digit should be used in the fourth decimal place.   When a final 5 is not in italic, follow the usual rule of rounding 0.5 to the nearest even integer.   Thus a final pair 55 becomes 6, 3$5$ becomes 3, 8$5$ becomes 9.

For a detailed table see *The Kelley Statistical Tables;* Table I gives values of $K_\epsilon$ (denoted by $x$) to 8 decimal places for values of $\epsilon$ (denoted by $q$) from 0 to 0.5 by steps of 0.0001.   Alternatively, see Fisher and Yates, *Statistical Tables;* Table I gives values of $K_\epsilon$ (denoted by $x$) to 6 decimal places for values of $\epsilon$ (denoted by $P/2$) from 0 to 0.495 by steps of 0.005; also Table IX gives values of $K_\epsilon + 5$ to 4 decimal places for values of $1 - \epsilon$ from 0 to 0.980 by steps of 0.001 and from 0.9800 to 0.9999 by steps of 0.0001.

### 4.3. Illustration

To illustrate the use of equations (10), (11), and (12), suppose that lots are considered acceptable if the percent defective is 15 or less and unacceptable if the percent defective is 30 or more, and suppose the maximum producer's and consumer's risks are to be 1 percent and 2 percent, respectively.   Then

$$p_1 = 0.15 \qquad \alpha = 0.01 \tag{13}$$
$$p_2 = 0.30 \qquad \beta = 0.02 \tag{14}$$

From Table 1.1, using only 4 decimal places, we find

$$K_1 = 1.0364 \qquad K_\alpha = 2.3263 \tag{15}$$
$$K_2 = 0.5244 \qquad K_\beta = 2.0537 \tag{16}$$

Substituting from (15) and (16) into (10) gives

$$k = \frac{2.3263 \times 0.5244 + 2.0537 \times 1.0364}{2.3263 + 2.0537} \tag{17}$$
$$= \frac{3.3484}{4.3800}$$
$$= 0.7645$$

Substituting from (15) and (16) into (11) gives

$$N = \frac{2 \times 4.3800^2 + 3.3484^2}{2 \times 0.5120^2} \tag{18}$$
$$= 94.6$$

The same value of $N$ results from substituting from (15), (16), and (17) into (12)

$$N = \frac{0.7645^2 + 2}{2} \left(\frac{4.3800}{0.5120}\right)^2 \tag{19}$$
$$= 1.2922 \times 73.1829$$
$$= 94.6$$

Thus, in this case a sample of 95 items would be drawn at random from the lot and a measurement of the characteristic being inspected would be made on each item of the sample. The sum of the measurements and the sum of their squares would be computed. These sums would be substituted into equations (4) and (6) to find the sample mean and standard deviation, $\bar{x}$ and $s$. Finally, $\bar{x} + 0.7645s$ would be computed; if it exceeded $U$ (the value of the quality characteristic above which an item is considered defective) the lot would be rejected, otherwise the lot would be accepted. (Actually, the abbreviated computational procedures described in Sec. 7 might be used.)

### 4.4. Adjustment of $N$

If the sample size given by (11) and (12) is too large, two new points may be selected in such a way that the OC curve will be less steep. For example, if $p_1$ and $p_2$ are kept at 15 and 30 percent but $\alpha$ and $\beta$ are doubled, to 2 and 4 percent, $N$ is reduced to 72; or if $\alpha$ and $\beta$ are kept at 1 and 2 percent but $p_1$ and $p_2$ are changed to 10 and 35 percent, $N$ becomes 32.

### 4.5. Adjustment of $k$ with $N$ Fixed

**4.5.1. Formula.**—If for some reason $N$ is definitely fixed, the OC curve can be made to pass through any one point desired, say

$$p = p_1 \qquad L_{p_1} = 1 - \alpha \tag{20}$$

by setting

$$k = \frac{K_1 \pm \sqrt{K_1^2 - ab}}{a} \tag{21}$$

where the plus sign is to be used if $\alpha > 0.5$ and the minus sign if $\alpha < 0.5$, and where

$$a = 1 - \frac{K_\alpha^2}{2(N-1)} \tag{22}$$

and

$$b = K_1^2 - \frac{K_\alpha^2}{N} \tag{23}$$

**4.5.2. Example.**—Suppose, for example, that a sample of 95 has been decided upon, and a sampling plan is desired such that the OC curve passes through the point defined by (14). Then (20) becomes

$$p_1 = 0.30 \qquad L_{p_1} = 0.02 \tag{24}$$

and from Table 1.1 we find

$$K_1 = 0.5244 \qquad K_\alpha = -2.0537 \tag{25}$$

Hence

$$a = 1 - \frac{4.2177}{188} \tag{26}$$

$$= 0.9776$$

$$b = 0.2750 - \frac{4.2177}{95} \tag{27}$$

$$= 0.2306$$

$$ab = 0.2254 \tag{28}$$

$$K_1^2 - ab = 0.2750 - 0.2254 \tag{29}$$

$$= 0.0496$$

$$\sqrt{K_1^2 - ab} = 0.2227 \tag{30}$$

$$k = \frac{0.5244 + 0.2227}{0.9776} \tag{31}$$

$$= 0.7642$$

This differs slightly from the value 0.7645 yielded by (17) for a sample of 95 and an OC curve passing through the points defined by (13) and (14). The value of $k$ yielded by (31) would be equal (except for differences due to rounding) to that yielded by (17) if in (22) and (23) both $N$ and $N - 1$ were taken as 94.6.

**4.5.3. A Special Case : Indifference Quality.**—If $N$ is fixed and $k$ is to be selected so that the OC curve will pass through one specified point, this point might reasonably be chosen at the indifference quality $p_0$. The indifference quality is the quality at which it is a matter of complete indifference whether the lot is accepted or rejected.* Lots of better

---

\* For a discussion of the indifference quality as a key to the selection of sampling-inspection plans, see Statistical Research Group, *Sampling Inspection*, Chap. 15, Sec. 2.2.1.3, p. 145.

quality should be accepted more often than rejected, lots of worse quality should be rejected more often than accepted, and lots of the indifference quality should be accepted and rejected equally frequently. Thus, $\alpha$ in (20) becomes 0.5, $K_\alpha$ becomes 0, and (21) reduces to $k = K_0$, the normal deviate exceeded with probability equal to the indifference quality.

If $k$ is determined from the indifference quality, it does not depend upon $N$, for the term in (22) and the term in (23) that involve $N$ drop out when $K_\alpha = 0$. Thus if $k$ is determined from the indifference quality, changes in $N$ will not change the value of $p$ for which acceptance and rejection are equally likely, nor will they require changes in $k$. Roughly speaking, changes in $N$ affect the steepness of the curve, while changes in $k$ affect its location with respect to the quality scale.

### 4.6. Approximations in the Plan

Three kinds of approximation are involved in using formulas (10), (11), and (12). First, there is the approximation involved in assuming the quality characteristic to be normally distributed. Second, certain mathematical approximations are involved in the derivation of the formulas (see Sec. 11.2). Third, the value of $N$ given by formulas (11) and (12) is not ordinarily an integer, so an approximation is involved in substituting an integer. The effect of the first approximation is discussed later (Sec. 9.4). Even if the quality characteristic were exactly normally distributed, the result of the second and third approximations would be that the true OC curve would not pass precisely through the points intended; but it would come sufficiently close for all practical purposes. In the present illustration, for example, taking $N = 95$ and $k = 0.7645$ results in a producer's risk of 1 percent at a percent defective of 15.04 instead of 15. The inaccuracies due to the second and third approximations tend to be larger for smaller samples, both because the effect of the mathematical approximations is larger and because the effect of rounding $N$ is, on the average, larger; this point is discussed further in Secs. 4.7 and 11.2.

### 4.7. Table of $N$ and $k$

Table 1.2 contains values of $N$ and $k$ for 151 combinations of $p_1$ and $p_2$. The values of $p_1$ in the table range from 0.001 to 0.05; the values of $p_2$ vary from about $\frac{4}{3}p_1$ or $\frac{3}{2}p_1$ up to $10p_1$ or more. The values of $\alpha$ are in all cases 0.05 and the values of $\beta$ are 0.10. The values of $N$ and $k$ have been computed by methods explained in Sec. 11.5 that do not involve the mathematical approximations used in formulas (10) and (11) or (12), so they may be slightly different from, and more accurate than, the values given by (10) and (11) or (12). This is especially true when $N$ is small, not only because of the mathematical approximations under-

TABLE 1.2

VALUES OF $N$ AND $k$ FOR VARIOUS COMBINATIONS OF $p_1$ AND $p_2$ WHEN $\alpha = 0.05$ AND
$\beta = 0.10$

$N$ is·the sample size; $k$ is the multiple of the standard deviation in the criterion
$x + ks \leq U$ for accepting the lot, where $U$ is the value above which the characteristic
being inspected is regarded as defective; $p_1$ is the acceptable quality level (AQL);
$p_2$ is the lot tolerance percent defective (LTPD); $\alpha = 0.05$ is the producer's risk; and
$\beta = 0.10$ is the consumer's risk. The true $p_1$ and true $p_2$ are the values at which, for
the corresponding $N$ and $k$, $\alpha = 0.05$ and $\beta = 0.10$; the difference between the true
and intended values of $p_1$ and $p_2$ arises from the fact that $N$ must be an integer.

This table has been computed with the aid of Table IV in Johnson and Welch,
"Applications of the Non-central $t$-Distribution." The entries for $N$, $k$, true $p_1$, and
true $p_2$ are therefore slightly more accurate than those given by formulas (10), (11)
or (12), and (42), respectively. See Sec. 11.5 on the construction of this table.

| Intended $p_1$ | Intended $p_2$ | $N$ | $k$ | True $p_1$ | True $p_2$ |
|---|---|---|---|---|---|
| 0.001 | 0.0015 | 3177 | 3.0214 | 0.0010 | 0.0015 |
| | 0.002 | 1032 | 2.9711 | 0.0010 | 0.0020 |
| | 0.0025 | 567 | 2.9311 | 0.0010 | 0.0025 |
| | 0.003 | 381 | 2.8979 | 0.0010 | 0.0030 |
| | 0.004 | 226 | 2.8441 | 0.0010 | 0.0040 |
| | 0.005 | 161 | 2.8018 | 0.0010 | 0.0050 |
| | 0.006 | 125 | 2.7660 | 0.0010 | 0.0061 |
| | 0.007 | 103 | 2.7359 | 0.0010 | 0.0071 |
| | 0.008 | 87 | 2.7153 | 0.0010 | 0.0080 |
| | 0.009 | 76 | 2.6832 | 0.0010 | 0.0092 |
| | 0.01 | 68 | 2.6725 | 0.0010 | 0.0099 |
| | 0.012 | 56 | 2.6220 | 0.0010 | 0.0122 |
| | 0.015 | 45 | 2.5730 | 0.0010 | 0.0153 |
| | 0.02 | 34 | 2.5208 | 0 0010 | 0.0201 |
| | 0.025 | 28 | 2.4712 | 0.0010 | 0.0250 |
| | 0.03 | 24 | 2.4392 | 0 0010 | 0.0294 |
| | 0.035 | 21 | 2.3930 | 0.0010 | 0.0351 |
| | 0.04 | 19 | 2.3745 | 0.0010 | 0.0389 |
| | 0.05 | 16 | 2.3059 | 0.0010 | 0.0494 |
| | 0.06 | 14 | 2.2570 | 0.0010 | 0.0590 |
| 0.0025 | 0.004 | 1677 | 2.7199 | 0.0025 | 0.0040 |
| | 0.005 | 736 | 2.6772 | 0.0025 | 0.0050 |
| | 0.006 | 443 | 2.6413 | 0.0025 | 0.0060 |
| | 0.0075 | 268 | 2.5967 | 0.0025 | 0.0075 |
| | 0.01 | 158 | 2.5376 | 0.0025 | 0.0101 |
| | 0.012 | 118 | 2.4979 | 0.0025 | 0.0121 |
| | 0.015 | 86 | 2 4576 | 0.0025 | 0.0149 |
| | 0.02 | 59 | 2.3875 | 0.0025 | 0.0202 |
| | 0.025 | 46 | 2.3410 | 0.0025 | 0.0249 |
| | 0.03 | 37 | 2.2981 | 0.0025 | 0.0301 |
| | 0.035 | 32 | 2.2610 | 0.0025 | 0.0348 |
| | 0.04 | 28 | 2.2354 | 0.0025 | 0.0391 |
| | 0.05 | 22 | 2.1724 | 0.0025 | 0.0501 |
| | 0.06 | 19 | 2.1321 | 0.0025 | 0.0585 |

TABLE 1.2.—(*Continued*)

| Intended $p_1$ | Intended $p_2$ | $N$ | $k$ | True $p_1$ | True $p_2$ |
|---|---|---|---|---|---|
| 0.005 | 0.0075 | 1713 | 2.4952 | 0.0050 | 0.0075 |
| | 0.01 | 547 | 2.4357 | 0.0050 | 0.0100 |
| | 0.012 | 328 | 2.3967 | 0.0050 | 0.0120 |
| | 0.015 | 197 | 2.3482 | 0.0050 | 0.0151 |
| | 0.02 | 114 | 2.2824 | 0.0050 | 0.0202 |
| | 0.025 | 80 | 2.2297 | 0.0051 | 0.0253 |
| | 0.03 | 61 | 2.1886 | 0.0050 | 0.0302 |
| | 0.035 | 50 | 2.1546 | 0.0050 | 0.0349 |
| | 0.04 | 42 | 2.1272 | 0.0049 | 0.0395 |
| | 0.05 | 32 | 2.0723 | 0.0049 | 0.0492 |
| | 0.06 | 26 | 2.0265 | 0.0049 | 0.0587 |
| | 0.07 | 22 | 1.9874 | 0.0049 | 0.0680 |
| 0 0075 | 0.01 | 2907 | 2.3728 | 0.0075 | 0.0100 |
| | 0.012 | 1039 | 2.3339 | 0.0075 | 0.0120 |
| | 0.015 | 451 | 2.2851 | 0.0075 | 0.0150 |
| | 0.02 | 208 | 2.2198 | 0.0075 | 0.0201 |
| | 0.025 | 130 | 2.1700 | 0.0075 | 0.0250 |
| | 0.03 | 93 | 2.1274 | 0.0075 | 0.0299 |
| | 0.035 | 72 | 2.0891 | 0.0075 | 0.0350 |
| | 0.04 | 58 | 2.0555 | 0.0075 | 0.0401 |
| | 0.05 | 42 | 1.9996 | 0.0075 | 0.0500 |
| | 0.06 | 33 | 1.9501 | 0.0075 | 0.0601 |
| | 0.07 | 27 | 1.9093 | 0.0075 | 0.0700 |
| | 0.08 | 23 | 1.8714 | 0.0075 | 0.0800 |
| 0.01 | 0.015 | 1231 | 2.2385 | 0.0100 | 0.0150 |
| | 0.02 | 389 | 2.1733 | 0.0100 | 0.0200 |
| | 0.025 | 208 | 2.1205 | 0.0100 | 0.0251 |
| | 0.03 | 137 | 2.0761 | 0.0101 | 0.0302 |
| | 0.035 | 101 | 2.0372 | 0.0101 | 0.0353 |
| | 0.04 | 79 | 2.0041 | 0.0101 | 0.0403 |
| | 0.045 | 64 | 1.9776 | 0.0100 | 0.0451 |
| | 0.05 | 54 | 1.9527 | 0.0099 | 0.0499 |
| | 0.06 | 41 | 1.9007 | 0.0100 | 0.0602 |
| | 0.07 | 33 | 1.8561 | 0.0101 | 0.0704 |
| | 0.08 | 28 | 1.8271 | 0.0099 | 0.0787 |
| | 0.09 | 24 | 1.7974 | 0.0098 | 0.0880 |
| | 0.10 | 21 | 1.7643 | 0.0099 | 0.0979 |

TABLE 1.2.—(Continued)

| Intended $p_1$ | Intended $p_2$ | $N$ | $k$ | True $p_1$ | True $p_2$ |
|---|---|---|---|---|---|
| 0.015 | 0.02 | 2035 | 2.1047 | 0.0150 | 0.0200 |
| | 0.025 | 603 | 2.0520 | 0.0150 | 0.0250 |
| | 0.03 | 310 | 2.0077 | 0.0150 | 0.0301 |
| | 0.035 | 197 | 1.9688 | 0.0151 | 0.0352 |
| | 0.04 | 141 | 1.9348 | 0.0151 | 0.0402 |
| | 0.045 | 108 | 1.9033 | 0.0151 | 0.0453 |
| | 0.05 | 87 | 1.8814 | 0.0149 | 0.0497 |
| | 0.06 | 61 | 1.8271 | 0.0151 | 0.0604 |
| | 0.07 | 47 | 1.7864 | 0.0151 | 0.0701 |
| | 0.08 | 38 | 1.7468 | 0.0152 | 0.0801 |
| | 0.09 | 32 | 1.7223 | 0.0148 | 0.0884 |
| | 0.10 | 27 | 1.6856 | 0.0149 | 0.0994 |
| | 0.11 | 24 | 1.6548 | 0.0152 | 0.1088 |
| | 0.12 | 21 | 1.6330 | 0.0149 | 0.1183 |
| | 0.13 | 19 | 1.6002 | 0.0153 | 0.1288 |
| | 0.14 | 17 | 1.5767 | 0.0152 | 0.1392 |
| | 0.15 | 15 | 1.5528 | 0.0148 | 0.1514 |
| 0.02 | 0.03 | 835 | 1.9565 | 0.0200 | 0.0300 |
| | 0.035 | 417 | 1.9179 | 0.0200 | 0.0351 |
| | 0.04 | 260 | 1.8836 | 0.0201 | 0.0401 |
| | 0.045 | 182 | 1.8523 | 0.0201 | 0.0452 |
| | 0.05 | 138 | 1.8247 | 0.0201 | 0.0501 |
| | 0.06 | 90 | 1.7680 | 0.0205 | 0.0609 |
| | 0.07 | 65 | 1.7364 | 0.0199 | 0.0697 |
| | 0.08 | 50 | 1.6971 | 0.0199 | 0.0800 |
| | 0.09 | 41 | 1.6617 | 0.0200 | 0.0898 |
| | 0.10 | 34 | 1.6290 | 0.0200 | 0.1001 |
| | 0.11 | 29 | 1.5968 | 0.0201 | 0.1107 |
| | 0.12 | 26 | 1.5722 | 0.0203 | 0.1191 |
| | 0.13 | 23 | 1.5473 | 0.0203 | 0.1287 |
| | 0.15 | 18 | 1.5037 | 0.0196 | 0.1495 |
| | 0.17 | 15 | 1.4623 | 0.0195 | 0.1692 |
| | 0.20 | 12 | 1.4041 | 0.0196 | 0.1986 |
| 0.03 | 0.04 | 1334 | 1.8077 | 0.0300 | 0.0400 |
| | 0.045 | 643 | 1.7766 | 0.0300 | 0.0451 |
| | 0.05 | 390 | 1.7483 | 0.0300 | 0.0501 |
| | 0.06 | 198 | 1.6979 | 0.0301 | 0.0601 |
| | 0.07 | 125 | 1.6539 | 0.0302 | 0.0702 |
| | 0.08 | 88 | 1.6083 | 0.0306 | 0.0812 |
| | 0.09 | 67 | 1.5844 | 0.0298 | 0.0896 |
| | 0.10 | 53 | 1.5517 | 0.0298 | 0.0998 |
| | 0.11 | 44 | 1.5236 | 0.0298 | 0.1092 |
| | 0.12 | 37 | 1.4955 | 0.0298 | 0.1194 |
| | 0.13 | 32 | 1.4704 | 0.0298 | 0.1290 |
| | 0.15 | 25 | 1.4224 | 0.0299 | 0.1485 |
| | 0.20 | 15 | 1.3149 | 0.0297 | 0.2013 |
| | 0.25 | 11 | 1.2281 | 0.0305 | 0.2477 |
| | 0.30 | 8 | 1.1553 | 0.0293 | 0.3007 |

TABLE 1.2.—(*Continued*)

| Intended $p_1$ | Intended $p_2$ | $N$ | $k$ | True $p_1$ | True $p_2$ |
|---|---|---|---|---|---|
| 0.04 | 0.06 | 524 | 1.6406 | 0.0400 | 0.0601 |
|      | 0.07 | 259 | 1.5964 | 0.0401 | 0.0701 |
|      | 0.08 | 160 | 1.5569 | 0.0402 | 0.0802 |
|      | 0.09 | 111 | 1.5209 | 0.0402 | 0.0903 |
|      | 0.10 | 83  | 1.4920 | 0.0399 | 0.0997 |
|      | 0.11 | 65  | 1.4621 | 0.0399 | 0.1098 |
|      | 0.12 | 53  | 1.4348 | 0.0398 | 0.1196 |
|      | 0.13 | 45  | 1.4112 | 0.0398 | 0.1286 |
|      | 0.14 | 38  | 1.3793 | 0.0403 | 0.1402 |
|      | 0.15 | 33  | 1.3565 | 0.0402 | 0.1500 |
|      | 0.17 | 26  | 1.3195 | 0.0397 | 0.1681 |
|      | 0.20 | 19  | 1.2585 | 0.0396 | 0.1988 |
|      | 0.25 | 13  | 1.1647 | 0.0405 | 0.2491 |
|      | 0.30 | 9   | 1.0967 | 0.0378 | 0.3027 |
|      | 0.35 | 7   | 1.0274 | 0.0376 | 0.3524 |
|      | 0.40 | 6   | 0.9480 | 0.0412 | 0.3963 |
| 0.05 | 0.07 | 661 | 1.5499 | 0.0500 | 0.0701 |
|      | 0.08 | 320 | 1.5102 | 0.0501 | 0.0801 |
|      | 0.09 | 194 | 1.4740 | 0.0501 | 0.0902 |
|      | 0.10 | 133 | 1.4408 | 0.0502 | 0.1004 |
|      | 0.11 | 99  | 1.4096 | 0.0504 | 0.1104 |
|      | 0.12 | 77  | 1.3826 | 0.0503 | 0.1202 |
|      | 0.13 | 62  | 1.3567 | 0.0502 | 0.1302 |
|      | 0.14 | 52  | 1.3365 | 0.0498 | 0.1389 |
|      | 0.15 | 44  | 1.3129 | 0.0498 | 0.1488 |
|      | 0.16 | 38  | 1.2857 | 0.0503 | 0.1597 |
|      | 0.17 | 33  | 1.2636 | 0.0503 | 0.1700 |
|      | 0.20 | 24  | 1.2047 | 0.0506 | 0.1985 |
|      | 0.25 | 15  | 1.1189 | 0.0494 | 0.2499 |
|      | 0.30 | 11  | 1.0386 | 0.0503 | 0.2981 |
|      | 0.35 | 8   | 0.9633 | 0.0493 | 0.3537 |
|      | 0.40 | 6   | 0.9162 | 0.0448 | 0.4052 |

lying (10) and (11) or (12), but also because of the rounding of $N$. For example, if $p_1 = 0.04$, $p_2 = 0.40$, $\alpha = 0.05$, and $\beta = 0.10$, Table 1.2 shows $N = 6$ and $k = 0.9480$. The table also shows that for these values of $N$ and $k$ the true values of $p_1$ and $p_2$, which, because $N$ must be an integer, cannot be made precisely equal to the intended values, are 0.0412 and 0.3963. Formulas (10) and (11) yield for this example $k = 0.9091$ and $N = 5.4$, which is rounded upward to $N = 6$; for $N = 6$ and $k = 0.9091$ the true $p_1$ and $p_2$ are 0.0457 and 0.4073. If $k$ is adjusted by formula (21), taking $N = 6$, $p = 0.04$, and $\alpha = 0.05$, it becomes 0.9247. For this adjusted value of $k$ the true $p_1$ and $p_2$ are 0.0439 and

0.4029. Thus, even for this case where $N$ is only 6, it is questionable whether the greater accuracy of Table 1.2 or of the adjusted $k$, as compared with formulas (10) and (11) or (12), is of substantial value for most practical purposes; and for larger values of $N$ the differences are even less important.

## 5. AMOUNT OF INSPECTION

### 5.1. Comparison with Attribute Inspection

If each item were classified simply defective or nondefective and the acceptability of the lot judged by the number of defectives in a single sample, a sample of 145 would be required for an OC curve passing through the points defined by (13) and (14).* If the sampling were done sequentially by attributes, the average number of items inspected would be 20 for lots with no defectives, 63 for lots 15 percent defective, 133 for lots 22 percent defective, 61 for lots 30 percent defective, and 7 for lots 100 percent defective.†

TABLE 1.3

SAMPLE SIZES FOR EIGHT PAIRS OF SINGLE-SAMPLING PLANS BASED ON ATTRIBUTES AND VARIABLES

$\alpha = 0.05$, $\beta = 0.10$ for all plans
$N_a$ = sample size for attributes
$N_t$ = sample size for variables

| Plan number | $p_1$ | $p_2$ | $N_t$ | $N_a$ | $N_a/N_t$ | Saving $(N_a - N_t)/N_a$ percent |
|---|---|---|---|---|---|---|
| 1 | 0.001 | 0.002 | 1,032 | 12,476 | 12.09 | 91.7 |
| 2 | 0.001 | 0.06 | 14 | 46 | 3.29 | 69.6 |
| 3 | 0.01 | 0.015 | 1,231 | 4,185 | 3.40 | 70.6 |
| 4 | 0.01 | 0.10 | 21 | 44 | 2.10 | 52.3 |
| 5 | 0.05 | 0.07 | 661 | 1,199 | 1.81 | 44.9 |
| 6 | 0.05 | 0.40 | 6 | 10 | 1.67 | 40.0 |
| 7 | 0.15 | 0.20 | 323 | 493 | 1.53 | 34.5 |
| 8 | 0.15 | 0.40 | 17 | 26 | 1.53 | 34.6 |

For attributes the sample sizes have been computed as explained in Chap. 7. For variables they have been computed by the method described in Sec. 11.5 of this chapter, which sometimes gives values slightly different from, and slightly more accurate than, those given by formulas (11) and (12).

* See Chap. 7, Sec. 3, equation (21) and first footnote on p. 260.

† See Statistical Research Group, *Sequential Analysis*, Sec. 2 (second revision), where the example defined by equations (13) and (14) above is treated in some detail. Sequential analysis also can be used with variables (see Sec. 11.6 below) and will result in a reduction in the average amount of inspection similar to that with attributes —roughly 40 to 60 percent (see *Sequential Analysis*, Appendix A, p. A.03).

The savings in the number of items inspected by using variables instead of attributes depends partly upon the steepness of the OC curve and partly upon its location. The steeper the OC curve, the greater the saving fr·m using variables. The better the lot qualities involved (that is, the smaller the percentages defective), the greater the saving. Eight comparisons of attribute and variables plans are shown in Table 1.3. In each case, $\alpha = 0.05$ and $\beta = 0.10$. $N_a$ and $N_t$ represent the sample sizes for single-sampling plans based on attributes and variables, respectively.

## 5.2. Comparison with Variables Inspection When the Standard Deviation Is Known

If the true standard deviation of the characteristic being inspected is given, the proportion defective depends only upon the mean, being $p_1$ or less if the mean is $U - K_1\sigma$ or less, and $p_2$ or more if the mean is $U - K_2\sigma$ or more.* If the standard deviation were known, a sample of 73 would be required for an OC curve passing through the points defined by (13) and (14). The reduction of 22 observations (23 percent) from the 95 needed when the standard deviation is not known thus represents

TABLE 1.4

$$\alpha = 0.05, \qquad \beta = 0.10 \text{ for all plans}$$
$N_t$ = sample size with standard deviation unknown
$N_m$ = sample size with standard deviation known

| Plan number | $p_1$ | $p_2$ | $N_t$ | $N_m$ | $N_t/N_m$ | Saving $(N_t - N_m)/N_t$ percent |
|---|---|---|---|---|---|---|
| 1 | 0.001 | 0.002 | 1,032 | 190 | 5.43 | 81.6 |
| 2 | 0.001 | 0 06 | 14 | 4 | 3.50 | 71.4 |
| 3 | 0.01 | 0.015 | 1,231 | 351 | 3.51 | 71.5 |
| 4 | 0.01 | 0.10 | 21 | 8 | 2.62 | 61.9 |
| 5 | 0.05 | 0.07 | 661 | 300 | 2.20 | 54.6 |
| 6 | 0.05 | 0.40 | 6 | 4 | 1.50 | 33.3 |
| 7 | 0.15 | 0.20 | 323 | 226 | 1.43 | 30.0 |
| 8 | 0.15 | 0.40 | 17 | 14 | 1.21 | 17.6 |

For the known standard deviation, the sample sizes are given by $1/a$, where $a$ is defined by equation (137), Sec. 11.1 of this chapter. For the unknown standard deviation they have been computed by the method described in Sec. 11.5 of this chapter, which sometimes gives values slightly different from, and slightly more accurate than, those given by formulas (11) and (12).

the savings in inspection yielded by knowledge of the standard deviation.†
This saving, like that yielded by sufficient knowledge of the form of

* The use of variables in inspection for proportion defective when the standard deviation is known is discussed by Romig, *Allowable Average in Sampling Inspection.*

† Incidentally, 22 is the number of observations required for an OC curve satisfying (13) and (14) if the true mean is known and only the standard deviation need be

distribution to permit the use of variables instead of attributes, varies with the OC curve. Comparisons between $N_t$, the sample size for the variables plans with unknown mean and standard deviation, and $N_m$, the sample size with known standard deviation, are shown in Table 1.4 for eight plans defined by the same pairs of OC points used in Table 1.3. In general, the percentage saving, computed before $N_t$ and $N_m$ are rounded to integers, is $100k^2/(k^2 + 2)$, but for small samples it may differ from this on account of rounding.

## 6. OPERATING-CHARACTERISTIC CURVE

### 6.1. Introduction

The values of $k$ and $N$ may be obtained from (10) and (11) or (12) so that the OC curve will pass through two specified points. They may be obtained by selecting $N$ arbitrarily and computing $k$ from (21) so that the OC curve will pass through one specified point. They may both be determined arbitrarily. Regardless of how they are chosen, their effects must be evaluated in terms of the OC curve to which they give rise. How will they operate with lots of every quality that might conceivably be offered? If a lot is good, what is the risk that the sample from it will fail to meet the acceptance requirement? If a lot is bad, what is the risk that the sample will satisfy the acceptance criterion? The OC curve shows, for every possible lot quality $p$, the probability $L_p$ that a lot of that quality will, if submitted, be accepted on the basis of the sampling plan.

### 6.2. Five-point OC Curve

Three points on the OC curve are easily obtained: (1) When there are no defectives in the lot, acceptance is certain; (2) when the lot contains only defectives, rejection is certain; (3) when the proportion of defectives in the lot is $p_0$, such that $K_{p_0} = k$, acceptance and rejection are equally likely. Thus three points on the OC curve are given by

$$p = 0 \qquad L_0 = 1 \qquad\qquad (32)$$
$$p = p_0 \qquad L_{p_0} = 0.5 \qquad\qquad (33)$$
$$p = 1 \qquad L_1 = 0 \qquad\qquad (34)$$

used in testing percent defective. In general, if $N_t$ is the number of observations required when neither the mean nor the standard deviation is known, $N_m$ the number required when the standard deviation is known, and $N_s$ the number required when the mean is known, the relation $N_t = N_m + N_s$ will hold; in fact, $N_m = 2N_t/(k^2 + 2)$ and $N_s = k^2 N_t/(k^2 + 2)$. See Chap. 8 for methods of determining the sample size $N_s$ when inspection is based on the standard deviation; the sample size $N_m$, when inspection is based on the mean, can be determined from equation (148) in Sec. 11.1 of this chapter: it is $1/a$, where $a$ is defined by equation (137).

where $p_0$ is the probability that a normal deviate will exceed $k$. If $k$ and $N$ have been determined by (10) and (11) or (12), the OC curve will also pass through the two points

$$p = p_1 \qquad L_{p_1} = 1 - \alpha \qquad (35)$$
$$p = p_2 \qquad L_{p_2} = \beta \qquad (36)$$

A sketch of the OC curve made from these five points is sufficient for many purposes. For the illustrative example of Sec. 4.3, for example, we have

$$p = 0 \qquad\qquad L_0 = 1 \qquad (37)$$
$$p_1 = 0.15 \qquad\quad L_{0.15} = 0.99 \qquad (38)$$
$$p_0 = 0.2223 \qquad L_{0.2223} = 0.5 \qquad (39)$$
$$p_2 = 0.30 \qquad\quad L_{0.30} = 0.02 \qquad (40)$$
$$p = 1 \qquad\qquad L_1 = 0 \qquad (41)$$
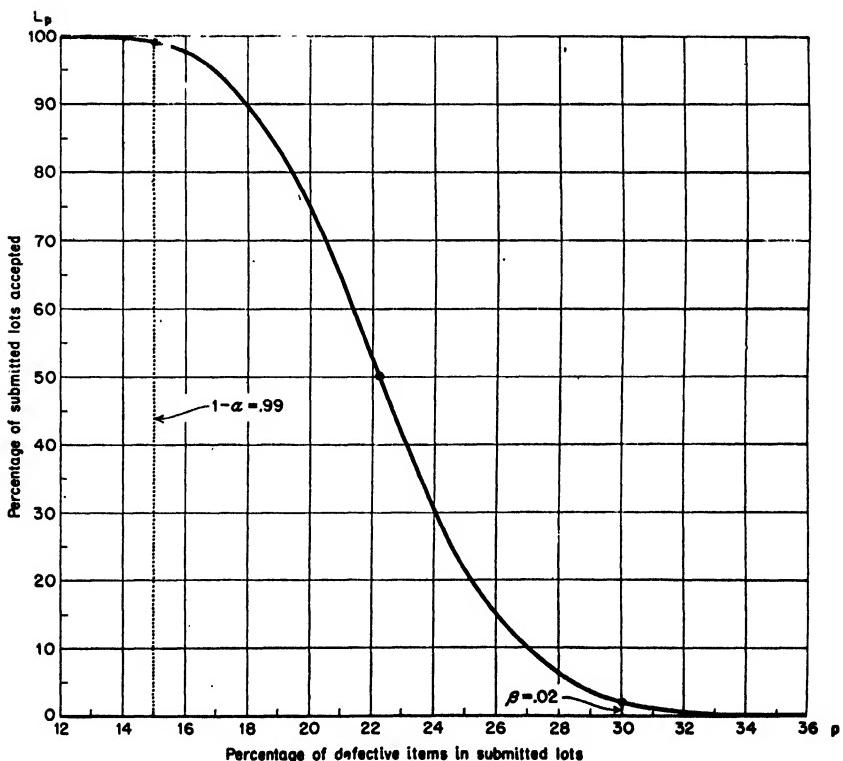


FIG. 1.4.—Operating-characteristic (*OC*) curve for a variables inspection plan for percent defective; $N = 95$, $k = 0.7645$. [*Data from Table 1.6, columns (6), (7), (8).*]

The central three of these five points are represented by dots in Fig. 1.4; the two extreme points are not shown because of the incompleteness of the $p$ scale.

## 6.3. Detailed OC Curve

As many points on the OC curve as desired may be found from the following formula, which expresses the normal deviate corresponding with the lot quality in terms of the normal deviate corresponding with any given probability of acceptance:

$$K_p = k - K_{L_p}\sqrt{\frac{1}{N} + \frac{k^2}{2(N-1)}} \tag{42}$$

Formula (42) yields the normal deviate corresponding with $p$. The value of $p$ can be found from Table 1.5, which shows the probabilities of exceeding various normal deviates.

If $p$ is to be computed for a number of values of $L_p$ in order to plot the entire OC curve, it is convenient while computing (42) to compute

$$K_{p'} = k + K_{L_p}\sqrt{\frac{1}{N} + \frac{k^2}{2(N-1)}} \tag{43}$$

If it is desired to find the probability of acceptance for a lot of given quality, (42) and (43) can be rewritten as

$$K_{L_p} = \frac{k - K_p}{\sqrt{\frac{1}{N} + \frac{k^2}{2(N-1)}}} \tag{44}$$

$$K_{L_{p'}} = -K_{L_p} \tag{45}$$

**6.3.1. Illustrative Computation for One Point.**—To illustrate the use of formulas (42) and (43), consider an acceptance-inspection plan defined by $N = 95$ and $k = 0.7645$. For what lot quality is the probability of rejection 1 percent?

We have

$$N = 95 \qquad k = 0.7645 \qquad K_{0.99} = -2.32635 \tag{46}$$

Then

$$\sqrt{\frac{1}{N} + \frac{k^2}{2(N-1)}} = 0.11677 \tag{47}$$

$$K_{L_p}\sqrt{\frac{1}{N} + \frac{k^2}{2(N-1)}} = -0.27165 \tag{48}$$

Substituting into (42),

$$K_p = 1.03615 \tag{49}$$

## TABLE 1.5
### PROBABILITIES CORRESPONDING WITH GIVEN NORMAL DEVIATES

Value of $\epsilon$ for various $K_\epsilon$, where $\dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{K_\epsilon}^{\infty} e^{-\frac{x^2}{2}} \, dx = \epsilon$
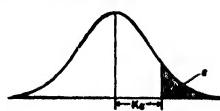
| $K_\epsilon$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 0.5000 | 0.4960 | 0.4920 | 0.4880 | 0.4840 | 0.4801 | 0.4761 | 0.4721 | 0.4681 | 0.4641 |
| 0.1 | 0.4602 | 0.4562 | 0.4522 | 0.4483 | 0.4443 | 0.4404 | 0.4364 | 0.4325 | 0.4286 | 0.4247 |
| 0.2 | 0.4207 | 0.4168 | 0.4129 | 0.4090 | 0.4052 | 0.4013 | 0.3974 | 0.3936 | 0.3897 | 0.3859 |
| 0.3 | 0.3821 | 0.3783 | 0.3745 | 0.3707 | 0.3669 | 0.3632 | 0.3594 | 0.3557 | 0.3520 | 0.3483 |
| 0.4 | 0.3446 | 0.3409 | 0.3372 | 0.3336 | 0.3300 | 0.3264 | 0.3228 | 0.3192 | 0.3156 | 0.3121 |
| 0.5 | 0.3085 | 0.3050 | 0.3015 | 0.2981 | 0.2946 | 0.2912 | 0.2877 | 0.2843 | 0.2810 | 0.2776 |
| 0.6 | 0.2743 | 0.2709 | 0.2676 | 0.2643 | 0.2611 | 0.2578 | 0.2546 | 0.2514 | 0.2483 | 0.2451 |
| 0.7 | 0.2420 | 0.2389 | 0.2358 | 0.2327 | 0.2296 | 0.2266 | 0.2236 | 0.2206 | 0.2177 | 0.2148 |
| 0.8 | 0.2119 | 0.2090 | 0.2061 | 0.2033 | 0.2005 | 0.1977 | 0.1949 | 0.1922 | 0.1894 | 0.1867 |
| 0.9 | 0.1841 | 0.1814 | 0.1788 | 0.1762 | 0.1736 | 0.1711 | 0.1685 | 0.1660 | 0.1635 | 0.1611 |
| 1.0 | 0.1587 | 0.1562 | 0.1539 | 0.1515 | 0.1492 | 0.1469 | 0.1446 | 0.1423 | 0.1401 | 0.1379 |
| 1.1 | 0.1357 | 0.1335 | 0.1314 | 0.1292 | 0.1271 | 0.1251 | 0.1230 | 0.1210 | 0.1190 | 0.1170 |
| 1.2 | 0.1151 | 0.1131 | 0.1112 | 0.1093 | 0.1075 | 0.1056 | 0.1038 | 0.1020 | 0.1003 | 0.0985 |
| 1.3 | 0.0968 | 0.0951 | 0.0934 | 0.0918 | 0.0901 | 0.0885 | 0.0869 | 0.0853 | 0.0838 | 0.0823 |
| 1.4 | 0.0808 | 0.0793 | 0.0778 | 0.0764 | 0.0749 | 0.0735 | 0.0721 | 0.0708 | 0.0694 | 0.0681 |
| 1.5 | 0.0668 | 0.0655 | 0.0643 | 0.0630 | 0.0618 | 0.0606 | 0.0594 | 0.0582 | 0.0571 | 0.0559 |
| 1.6 | 0.0548 | 0.0537 | 0.0526 | 0.0516 | 0.0505 | 0.0495 | 0.0485 | 0.0475 | 0.0465 | 0.0455 |
| 1.7 | 0.0446 | 0.0436 | 0.0427 | 0.0418 | 0.0409 | 0.0401 | 0.0392 | 0.0384 | 0.0375 | 0.0367 |
| 1.8 | 0.0359 | 0.0351 | 0.0344 | 0.0336 | 0.0329 | 0.0322 | 0.0314 | 0.0307 | 0.0301 | 0.0294 |
| 1.9 | 0.0287 | 0.0281 | 0.0274 | 0.0268 | 0.0262 | 0.0256 | 0.0250 | 0.0244 | 0.0239 | 0.0233 |
| 2.0 | 0.0228 | 0.0222 | 0.0217 | 0.0212 | 0.0207 | 0.0202 | 0.0197 | 0.0192 | 0.0188 | 0.0183 |
| 2.1 | 0.0179 | 0.0174 | 0.0170 | 0.0166 | 0.0162 | 0.0158 | 0.0154 | 0.0150 | 0.0146 | 0.0143 |
| 2.2 | 0.0139 | 0.0136 | 0.0132 | 0.0129 | 0.0125 | 0.0122 | 0.0119 | 0.0116 | 0.0113 | 0.0110 |
| 2.3 | 0.0107 | 0.0104 | 0.0102 | 0.0099 | 0.0096 | 0.0094 | 0.0091 | 0.0089 | 0.0087 | 0.0084 |
| 2.4 | 0.0082 | 0.0080 | 0.0078 | 0.0075 | 0.0073 | 0.0071 | 0.0069 | 0.0068 | 0.0066 | 0.0064 |
| 2.5 | 0.0062 | 0.0060 | 0.0059 | 0.0057 | 0.0055 | 0.0054 | 0.0052 | 0.0051 | 0.0049 | 0.0048 |
| 2.6 | 0.0047 | 0.0045 | 0.0044 | 0.0043 | 0.0041 | 0.0040 | 0.0039 | 0.0038 | 0.0037 | 0.0036 |
| 2.7 | 0.0035 | 0.0034 | 0.0033 | 0.0032 | 0.0031 | 0.0030 | 0.0029 | 0.0028 | 0.0027 | 0.0026 |
| 2.8 | 0.0026 | 0.0025 | 0.0024 | 0.0023 | 0.0023 | 0.0022 | 0.0021 | 0.0021 | 0.0020 | 0.0019 |
| 2.9 | 0.0019 | 0.0018 | 0.0018 | 0.0017 | 0.0016 | 0.0016 | 0.0015 | 0.0015 | 0.0014 | 0.0014 |
| 3.0 | 0.0013 | 0.0013 | 0.0013 | 0.0012 | 0.0012 | 0.0011 | 0.0011 | 0.0011 | 0.0010 | 0.0010 |

The entries in row $(K_\epsilon)$ are additional digits for the values of $K_\epsilon$ shown in the first column. Thus, the entry 0.0934 in column (2) and row (1.3) corresponds with $K_\epsilon = 1.32$; similarly, the entry 0.4840 in column (4) and row (0.0) corresponds with $K_\epsilon = 0.04$.

A *5* in italic indicates that, in rounding, the nearest *odd* digit should be used in the preceding decimal place. When a 5 is not in italic, follow the usual rule of rounding to the nearest even digit. Thus a final pair 55 becomes 6, 3*5* becomes 3, 8*5* becomes 9.

For a detailed table see New York Mathematical Tables Project *Tables of Probability Functions*. Table I gives values of $1 - 2\epsilon$ to 15 decimal places for values of $K_\epsilon$ (denoted by $x$) from 0 to 1 by steps of 0.0001, and for all values of $K_\epsilon$ above 1 by steps of 0.001.

Interpolating linearly in Table 1.5, we find $p = 15.01$ percent.* Similarly, on substituting into (43) we find $K_{p'} = 0.49285$; so $p = 31.11$ percent for $L_p = 0.01$.

**6.3.2. Work Sheet for Detailed OC Curve.**—When a number of points on the OC curve are needed, it is convenient to arrange the computations in a systematic way, as for example that shown in Table 1.6, which contains the computations for the OC curve of the example of Sec. 4.3.

<div align="center">

TABLE 1.6

WORK SHEET FOR OPERATING-CHARACTERISTIC CURVE

</div>

Basic quantities: $N = 95$; $\quad k = 0.7645$; $\quad S = \sqrt{\dfrac{1}{N} + \dfrac{k^2}{2(N-1)}} = 0.11677$

| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|-----|-----|-----|-----|-----|-----|-----|-----|
| $L_p$ | $K_{L_p}$ | $S \times (2)$ | $k - (3)$ | $k + (3)$ | $p$ | $p'$ | $1 - (1) = L_{p'}$ |
| 0.005 | 2.57583 | 0.30078 | 0.46372 | 1.06528 | 0.321 | 0.143 | 0.995 |
| 0.01 | 2.32635 | 0.27165 | 0.49285 | 1.03615 | 0.311 | 0.150 | 0.99 |
| 0.02 | 2.05375 | 0.23982 | 0.52468 | 1.00432 | 0.300 | 0.158 | 0.98 |
| 0.03 | 1.88079 | 0.21962 | 0.54488 | 0.98412 | 0.293 | 0.163 | 0.97 |
| 0.04 | 1.75069 | 0.20443 | 0.56007 | 0.96893 | 0.288 | 0.166 | 0.96 |
| 0.05 | 1.64485 | 0.19207 | 0.57243 | 0.95657 | 0.284 | 0.169 | 0.95 |
| 0.15 | 1.03643 | 0.12102 | 0.64348 | 0.88552 | 0.260 | 0.188 | 0.85 |
| 0.25 | 0.67449 | 0.07876 | 0.68574 | 0.84326 | 0.246 | 0.200 | 0.75 |
| 0.50 | 0 | 0 | 0.76450 | 0.76450 | 0.222 | 0.222 | 0.50 |

*First*, enter at the top of the sheet the three basic quantities $N$, $k$, and $S$, $S$ being defined by the left-hand side of equation (47).

*Second*, enter in column (1) various ordinates $L_p$ less than or equal to 0.5 for which the corresponding abscissas $p$ are to be computed.

*Third*, enter in column (2) values $K_{L_p}$ of the normal deviates corresponding with the values of $L_p$; these may be obtained from Table 1.1.

*Fourth*, enter in column (3) the products of $S$ by the normal deviates ·in column (2).

*Fifth*, subtract from $k$ the entries in column (3) and enter the differences in column (4).

*Sixth*, add to $k$ the entries in column (3) and enter the sums in column (5).

*Seventh*, enter in column (6) the probabilities corresponding with the normal deviates in column (4); these may be obtained from Table 1.5.

*Eighth*, enter in column (7) the probabilities corresponding with the normal deviates in column (5); these may be obtained from Table 1.5.

---

* The value 15.04 percent stated in Sec. 4.6 was obtained by an exact method, described in Sec. 11.5, which does not involve the mathematical approximations involved in formulas (42) and (43).

*Ninth*, subtract from 1 the values of $L_p$ shown in the first column and enter the differences in column (8).

*Tenth*, plot one-half the OC curve, using as abscissas the values of $p$ shown in column (6) and as ordinates the corresponding values of $L_p$ shown in column (1).

*Eleventh*, plot the other half of the OC curve, using as abscissas the values of $p'$ shown in column (7) and as ordinates the corresponding values of $L_{p'}$ shown in column (8).

The OC curve corresponding with Table 1.6 is shown in Fig. 1.4. The three dots represent the central three points of Sec. 6.2, formulas (38) to (40).

## 7. TECHNIQUES OF APPLYING THE PLAN

### 7.1 Introduction

Without calculating either the mean $\bar{x}$ or the standard deviation $s$, the computing can be arranged to show whether the acceptance criterion

$$\bar{x} + ks \leq U \tag{50}$$

has been met. This is done partly by rearranging formula (50), and partly by carrying out in advance certain steps in the computations that do not depend upon the data.

There is some question as to how much the computational simplifications add to the procedure. They are likely to be welcomed by those who consider the standard deviation "too complicated for practical purposes," and who would therefore reject (50) as an acceptance criterion if it were not emphasized that the standard deviation need not actually be computed. Those holding this view, however, often have had little or no practical experience in computing or using standard deviations, and what experience they have had may have been with inefficient computing techniques. In fact, statisticians do not generally realize how often standard deviations are computed by the naïve use of formula (5). The mean is calculated, then the $N$ differences between it and the observations are found and recorded; these are then squared and added, the sum divided by $N - 1$, and the square root taken. Not only is this process tedious and liable to error, but there is an inherent loss of accuracy from rounding the differences. The standard deviation should, of course, be computed from formula (6). With modern computing machines it can be calculated simultaneously with the calculation of the mean and with very little extra trouble. With the Monroe Model AA-1, for example, the first observation is set in the keyboard and the $\times$ and $=$ keys pressed in rapid succession; then the next observation is set on the keyboard and the same two keys pressed again; and so on. When this has been completed for all $N$ observations, which takes little more time and trouble than adding the

data, the quantities $\Sigma x$ and $\Sigma x^2$ needed for formula (6) are found in the upper and lower dials, respectively.*

Another reason computational simplifications are perhaps less important than they may at first appear is that the actual values of the mean and standard deviation, although not needed for acceptance purposes, are valuable for other purposes, as, for example, process control. In fact, the conditions under which inspection is based on variables instead of attributes are likely to be those where the cost of obtaining each observation is so high, because of destruction of test items, use of costly test equipment, and so forth, that the cost of analyzing the data is almost negligible in comparison.

## 7.2. Computational Procedure†

**7.2.1. Formulas.**—For any given plan of action, determined by a fixed $N$ and $k$, the left-hand side of formula (50) depends only upon the sum of the observations $\Sigma x$ and the sum of the squares of the observations $\Sigma x^2$. Let

$$X = \Sigma x \tag{51}$$
$$Y = \Sigma x^2 \tag{52}$$

These two sums may be computed easily and simultaneously on any automatic computing machine. If only an adding machine is available, $Y$ can be obtained with the aid of a table of squares, for example, Barlow's.

If $k$ is positive, it may be noted first that if

$$X > NU \tag{53}$$

the lot may be rejected without regard to $Y$. In that case $\bar{x}$ exceeds $U$; since $s$ must be positive, $\bar{x} + ks$ will still further exceed $U$.

Second, it may be noted that if

$$Y \leq \frac{U^2(N-1)}{k^2+1} \tag{54}$$

the lot may be accepted without regard to $X$. Equation (72) below is, however, more efficient than (54) in permitting immediate acceptance.

---

* This paragraph indicates only the general nature of the procedures for computing the standard deviation. Further simplifications are usually possible. Also, the use of the Monroe Model AA-1 as an illustration does not imply that this make or model is superior to other automatic calculating machines. Both the Fridén and Marchant companies make excellent machines. The Statistical Research Group attempted to evaluate various makes and models on the basis of about 660 machine-months of experience in its offices from 1942 to 1945. The evidence failed to show any one make or model to be best, although there were interesting differences. The terms "upper dials" and "lower dials" used in this chapter apply to Marchant and Monroe calculators but should be interchanged for Fridén calculators.

† For a discussion of the accuracy with which the measurements should be made, see Chap. 4.

Third, if neither inequality (53) nor (54) is satisfied, the lot is accepted if

$$Y \leq aX^2 - bX + c \tag{55}$$

where

$$a = \frac{Nk^2 + N - 1}{N^2 k^2} \tag{56}$$

$$b = \frac{2U(N - 1)}{Nk^2} \tag{57}$$

$$c = \frac{U^2(N - 1)}{k^2} \tag{58}$$

If inequality (55) is not fulfilled, the lot is rejected. The values of $a$, $b$, and $c$ given by (56), (57), and (58) do not depend upon the outcome of the inspection but only upon the sampling plan defined by $N$ and $k$ and upon the definition of a defective item as given by $U$. When a sample has been inspected, the values of $X$ and $Y$ are obtained from the observations and substituted into (55). Formula (78) below provides an additional condition, linear in $X$, that may be inserted between (54) and (55), which are, respectively, independent of and quadratic in $X$; formula (78), however, requires a preliminary estimate of the region in which the sample mean is likely to lie.

**7.2.2. Example.**—Consider a sampling plan in which $N = 95$ and $k = 0.7645$. Suppose that for the characteristic being inspected values in excess of $U = 35$ are regarded as defective.

From these values of $N$, $k$, and $U$ we have, from (56), (57), and (58),

$$a = 0.0283471 \tag{59}$$
$$b = 118.508 \tag{60}$$
$$c = 197,019 \tag{61}$$

Now suppose that a sample shows

$$X = \Sigma x = 2,872 \tag{62}$$
$$Y = \Sigma x^2 = 89,175 \tag{63}$$

First, it is noted that, since

$$X < 3,325 \tag{64}$$

inequality (53) is not fulfilled, so immediate rejection is not called for. Second, it is noted that, since

$$Y > 72,675 \tag{65}$$

immediate acceptance is not possible on the basis of (54). Applying (55), it is found that

$$Y < 90,482 \tag{66}$$

Therefore, the lot is acceptable.

In this example

$$\bar{x} = 30.232 \tag{67}$$
$$s = 5.000 \tag{68}$$
$$\bar{x} + ks = 34.05 \tag{69}$$

and, since this is less than 35, acceptance is indicated.

## 7.3. Graphical Procedure

**7.3.1. Original Data.**—The computing involved in substituting $X$ and $X^2$ into formula (55) may be avoided by the use of a chart. Values of $X$ may be plotted horizontally and values of $Y$ vertically, and a curve (parabola) drawn showing for each value of $X$ the minimum value of $Y$, $Y_R$, requiring rejection as determined from formula (55) using the equality sign. Then as soon as $X$ and $Y$ are determined from a sample, a point may be plotted on the chart. If the point falls on or below the parabola, the lot is accepted; if the point falls above the parabola, the lot is rejected.

Figure 1.5 shows such a curve for the example of Sec. 7.2.2, and the point corresponding with the sample of that example is shown.* Also shown in the figure is a lower boundary for the acceptance region. This lower boundary arises from the fact that, for any given sum of the observations $X$, it is algebraically impossible for the sum of the squares of the observations $Y$ to be less than a certain minimum; this minimum is the value $Y$ takes when all the observations are equal, namely,

$$Y_E = \frac{X^2}{N} \tag{70}$$

If a sample point should fall on the lower line, it would indicate that the standard deviation was zero. A sample point cannot fall below the lower line unless there is an error in the calculations.

For practical purposes only a part of the range of $X$ shown in Fig. 1.5 would be needed, for example, that enclosed in the rectangle. Confining the chart to this region, as in Fig. 1.6, permits the use of a larger scale and thus improves the accuracy of plotting.

**7.3.2. Adjusted Data.**—The graphical procedure of Sec. 7.3.1 is impractical when the observations are large. In such cases the range of variation in $Y$ as $X$ varies is large relative to the width of the acceptance band for a fixed $X$; when the $Y$ scale is contracted sufficiently to go on a sheet of reasonable size, the upper and lower boundaries of the acceptance region are pulled so close together that the chart becomes useless. A remedy for this is to replace the sum of the squares of the observations $Y$

---

* Formula (54) gives the minimum value of formula (55). Actually, it gives a value slightly below the minimum, because the minimum, which occurs when $X = UN(N-1)/(Nk^2 + N - 1)$, is $NU^2(N-1)/(Nk^2 + N - 1)$. If $Y$ is as small as (54), it is impossible for $X$ to be too large; this is illustrated in Fig. 1.5.

by the sum of the squares of adjusted observations $Y'$ where

$$Y' = \Sigma(x - A)^2 \tag{71}$$
$$= Y - 2AX + NA^2$$

and $A$ is an adjustment constant. The adjustment constant is, in effect, subtracted from each observation before squaring; actually the sum of squares of the adjusted observations, $Y'$, is obtained directly from $X$ and $Y$ by formula (71) without adjusting each observation.

In order to make the chart most useful, the adjustment constant $A$ should, if possible, be chosen near the mean of the characteristic being inspected. Usually there will be some evidence as to the general neighborhood in which the mean lies. If there is no evidence at all, it will often be suitable to choose $A$ somewhere in the interval between $U/(k^2 + 1)$ and $U$.

To compute $Y'$, $X$ and $Y$ are obtained simultaneously as described in Sec. 7.1, and $X$ is recorded. Then $X$ is copied to the keyboard and multiplied negatively by $2A$, which may be computed in advance. Next $NA^2$, which also should be computed in advance, is added. These two operations adjust the sum of squares, and $Y'$ may be copied from the lower dials. An alternative computing procedure, which may be useful if the same inspection plan, or at least the same adjustment constant, is to be used repeatedly, is to prepare a table showing $(x - A)^2$ for the various values of $x$ that might be observed. From this table $Y'$ can be found by simple addition.

The values of $X$ and $Y'$ determine whether the acceptance criterion has been met. If $X > NU$, immediate rejection may be made, according to formula (53). Instead of (54) for immediate acceptance, we have

$$Y' \leq \frac{(U - A)^2(N - 1)}{k^2 + 1} \tag{72}$$

If this is fulfilled the lot may be accepted immediately. Notice that $Y'$ may satisfy (72) even when $Y$ fails to satisfy (54). Otherwise, a chart similar to Fig. 1.6 may be drawn in which the upper boundary of the acceptance region is given by

$$Y'_R = aX^2 - (b + 2A)X + (c + NA^2) \tag{73}$$

where $a$, $b$, and $c$ are defined by (56), (57), and (58), and the lower boundary is given by

$$Y'_E = \frac{X^2}{N} - 2AX + NA^2 \tag{74}$$

Figure 1.7 shows such a chart for the example of Sec. 7.2.2, the adjustment constant being taken as $A = 30$. Thus, the equations of the
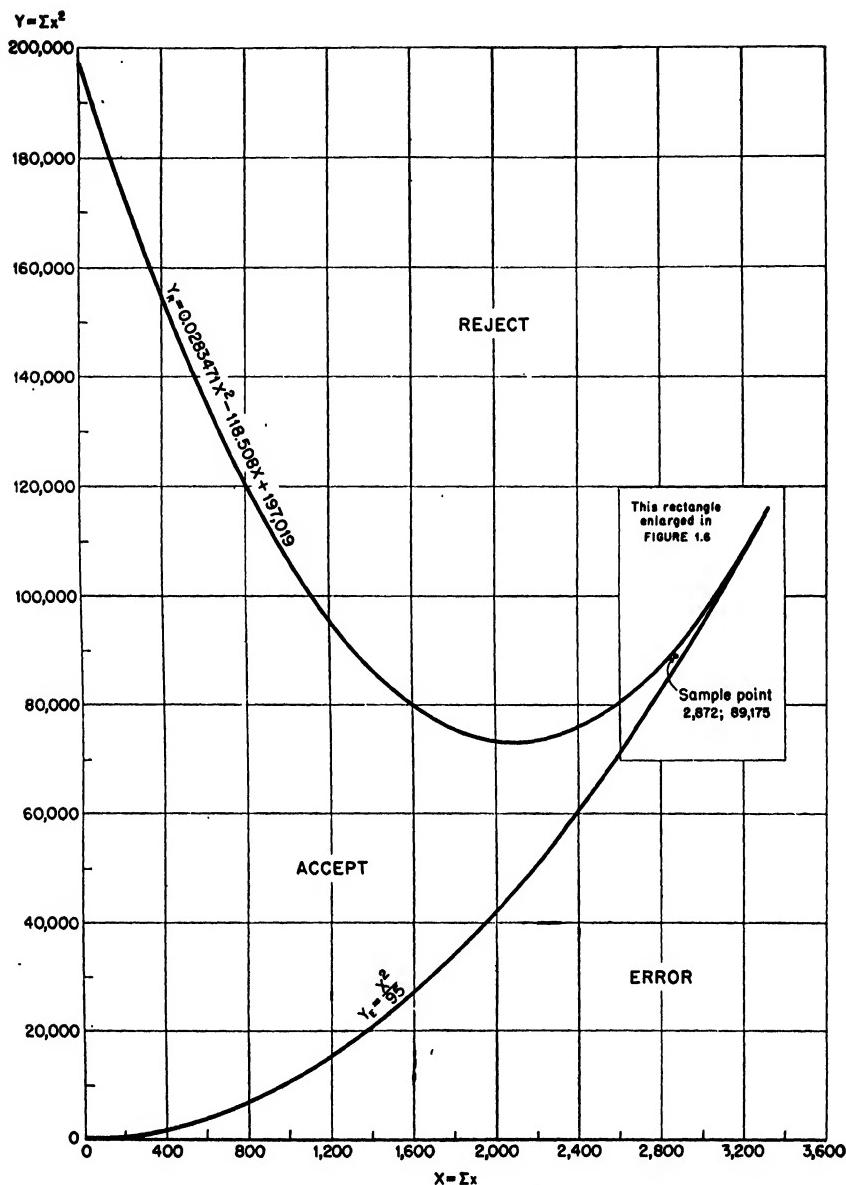
NOTE: The $Y_E$ curve should extend indefinitely, separating the error and rejection regions, but in a small figure this would make it difficult to identify the termination point of the acceptance region.

FIG. 1.5.—Graphical procedure for determining from $\Sigma x$ and $\Sigma x^2$ whether $\bar{x} + ks \leq U$; $N = 95$, $k = 0.7645$, $U = 35$.
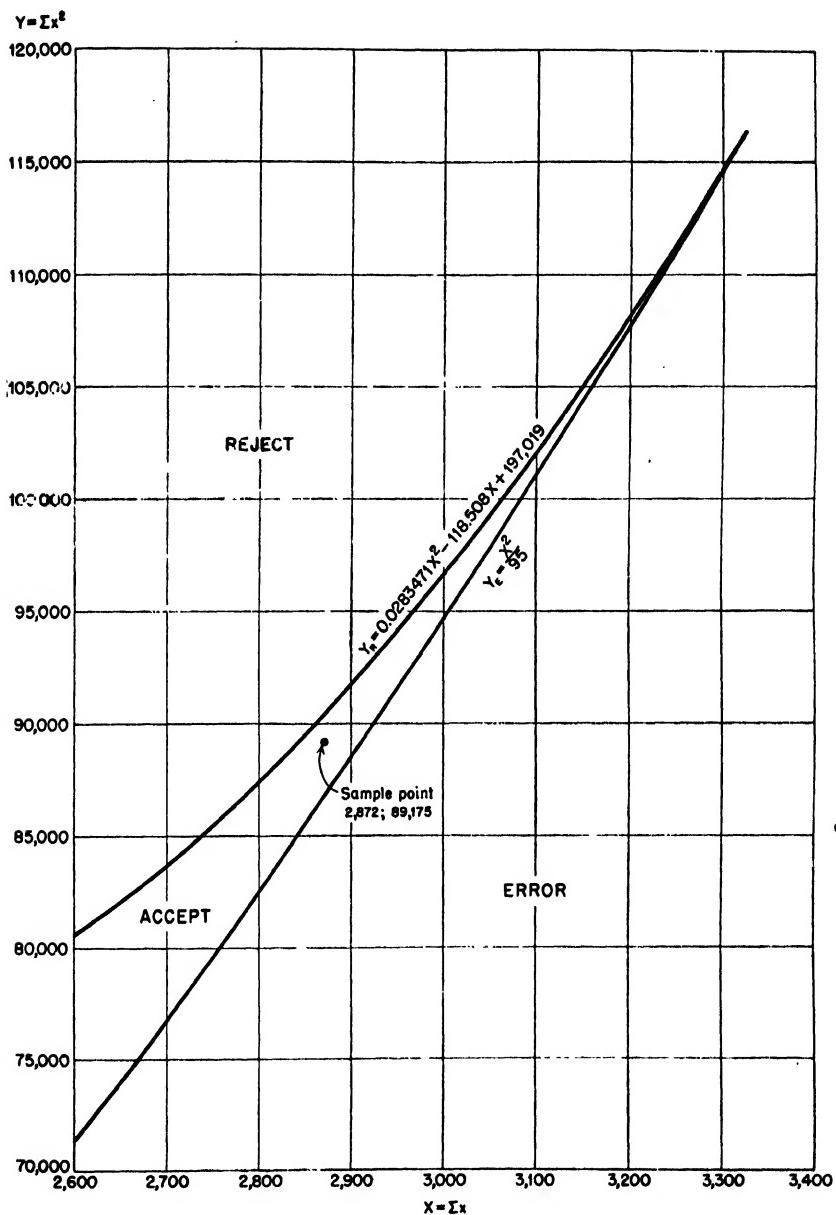
FIG. 1.6.—Enlarged segment of computational chart in terms of original data (enlargement of area indicated in Fig 1.5).

NOTE: See note to Fig. 1.5.
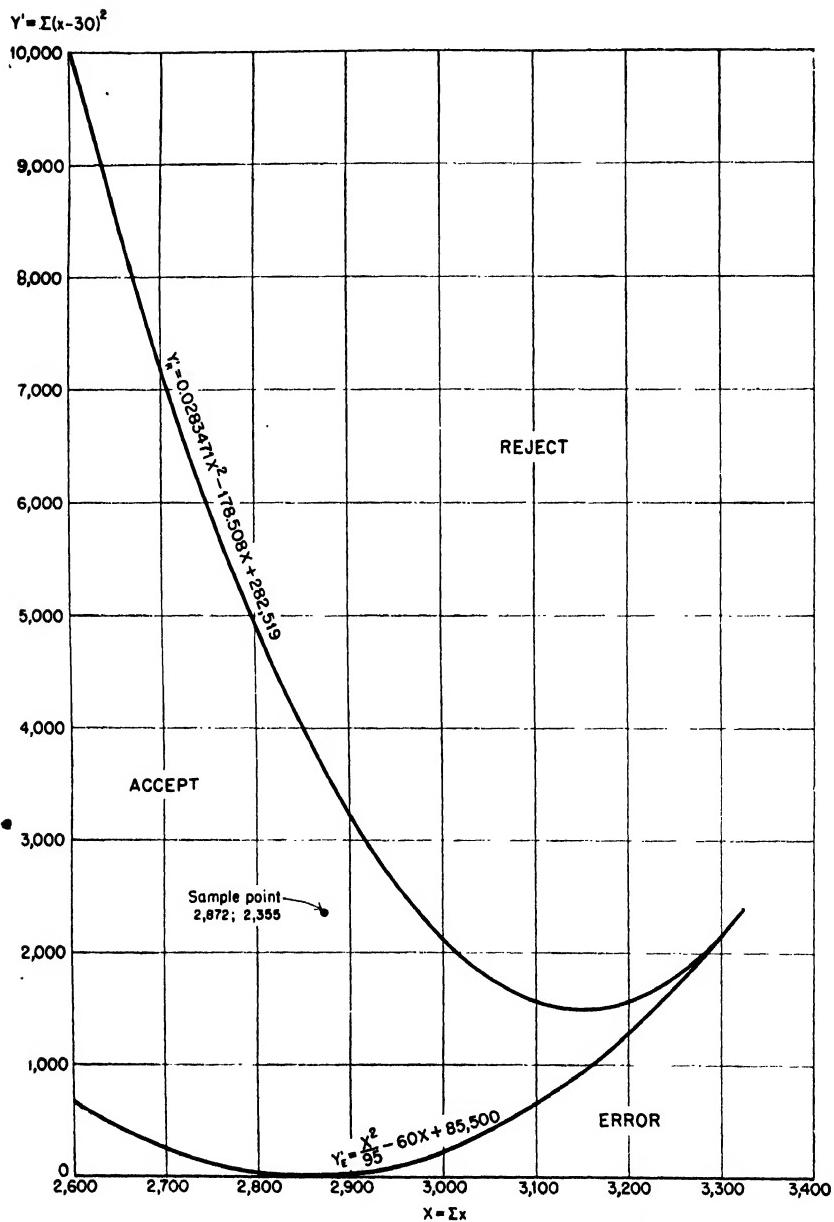
Note: See note to Fig. 1.5.

Fig. 1.7.—Graphical procedure for determining from $\Sigma x$ and $\Sigma(x - A)^2$ whether $\bar{x} + ks \leq U$; $N = 95$, $k = 0.7645$, $U = 35$, $A = 30$.

upper and lower boundaries of the acceptance region are

$$Y_R' = 0.0283471X^2 - 178.508X + 282.519 \tag{75}$$
$$Y_E' = 0.0105263X^2 - 60X + 85.500 \tag{76}$$

Since $2A = 60$ and $NA^2 = 85,500$, the value of $Y'$ in the sample is, by formula (71), 2,355. This does not satisfy inequality (72), which would require that $Y'$ be 1,483 or less; but when the point $X = 2,872$, $Y' = 2,355$ is plotted in Fig. 1.7, it is seen to be within the acceptance region.

It is stated in Sec. 7.2.1 that (72) is more effective than (54) in permitting immediate acceptance. In the present example, (72) requires that $Y' < 1,483$, which (since $X := 2,872$, is equivalent to requiring that $Y \le 88,303$, whereas (54) requires that $Y \le 72,675$. Thus values of $Y$ between 72,675 and 88,303 will lead to immediate acceptance by (72) but not by (54). From the point of view of permitting acceptance on the basis of $Y'$, rather than the point of view of making the chart usable, very nearly the more efficient form of (72) is obtained by setting

$$A = \frac{(1 + k^2)\mu' - U}{k^2} \tag{77}$$

where $\mu'$ is the best estimate of $\bar{x}$ available before the sample is inspected. In the present example, if $\mu'$ is taken as 30, formula (77) shows $A$ about 21.5. Taking $A = 21$, formula (72) becomes $Y' \le 11,628$, and since the sample value of $Y'$ is 10,446, immediate acceptance is possible. The value of $A$ given by (77) may be made the basis of a condition for acceptance in terms of $Y$; intermediate between (54) and (55), which are, respectively, constant and quadratic in $X$, the linear condition for acceptance may be inserted

$$Y \le c' - b'X \tag{78}$$

where

$$b' = b - 2aN\mu' \tag{79}$$
$$c' = c - aN^2\mu'^2 \tag{80}$$

and $a$, $b$, and $c$ are defined by (56), (57), and (58). Taking $\mu' = 30$ in the illustrative example, (78) becomes $Y \le 90,468$, which leads to acceptance. If the value of $\mu'$ happens to coincide with $\bar{x}$, (78) becomes equivalent to (55).

## 8. REDUCED AND TIGHTENED INSPECTION

### 8.1. Purpose of Reducing or Tightening Inspection

It is emphasized in Secs. 3.4 and 6.1 that the OC curve shows the conditional probability that *if* a submitted lot has a given quality $p$ it will be accepted. It does not show the probability that a lot of quality $p$ will be accepted. This latter probability cannot be controlled

by the sampling plan alone, for it depends upon the probability that a lot of quality $p$ will be submitted.

In judging the value of the protection afforded by a sampling plan, information about the quality of submitted material is important. The fact that a plan rejects 90 percent of submitted lots having a certain unacceptable quality may be of little importance if a great many of the submitted lots are of that quality; in that case many of the lots passed by the plan will be unacceptable, even though most of the submitted lots are rejected. The fact that the plan rejects 90 percent of submitted lots of an unacceptable quality $p$ is also unimportant if submitted lots rarely have that quality, for then very little such material will be accepted, even though most submitted lots are passed. In the first case, when bad material is submitted very often, the protection given by the sampling plan may be inadequate and require tightening; but in the second case, when bad material is rarely submitted, the protection may be superfluous and reduction may be permissible.*

### 8.2. Criteria for Reducing or Tightening Inspection

**8.2.1. Process Average Percent Defective.**—The principal criterion for determining whether reduced inspection is permissible or tightened inspection necessary is the process average percent defective, usually called simply the process average or PA, which is an estimate of the percent defective computed from a number of lots.

This estimate is best stated as a confidence interval, that is, an interval within which it may be asserted with, say, 95 percent confidence that the true process average lies. If the estimate were stated as a single figure, it would invariably differ from the true process average because of sampling fluctuations; and the question would therefore arise as to whether the difference between the estimated process average and the acceptable quality level could reasonably be ascribed to sampling fluctuations, or whether on the other hand it indicated a real discrepancy between the true process average and the acceptable quality level. A confidence interval answers such questions. It is constructed from the sample data in such a way that 95 percent, or any other desired proportion $\gamma$, of the samples will lead to confidence intervals that overlap the true process average.

For the purpose of computing the process average, all data in the samples from a number $h$ of consecutive lots are thrown together into one large sample. The mean $\bar{x}$ and the standard deviation $s$ of this combined sample are computed from formulas (4) and (6). Alternatively, they

---

* For a general discussion of the principles and methods of tightening and reducing inspection, see Statistical Research Group, *Sampling Inspection*. The discussion there relates to inspection by attributes, but, insofar as it runs in terms of the OC curve and the amount of inspection, it is also applicable to variables.

may be computed from the means $\bar{x}_i$, the standard deviations $s_i$, and the sample sizes $N_i$ of the individual samples by the following formulas:

$$\bar{x} = \frac{N_1\bar{x}_1 + N_2\bar{x}_2 + \cdots + N_h\bar{x}_h}{N_1 + N_2 + \cdots + N_h} = \frac{\Sigma N_i\bar{x}_i}{N} \qquad (81)$$

$$s^2 = \frac{\begin{array}{c}(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2 + \cdots + (N_h - 1)s_h^2 \\ + N_1(\bar{x}_1 - \bar{x})^2 + N_2(\bar{x}_2 - \bar{x})^2 + \cdots + N_h(\bar{x}_h - \bar{x})^2\end{array}}{N_1 + N_2 + \cdots + N_h - 1}$$

$$(82)$$

$$= \frac{\Sigma(N_i - 1)s_i^2 + \Sigma N_i\bar{x}_i^2 - \dfrac{(\Sigma N_i\bar{x}_i)^2}{N}}{N - 1}$$

where

$$N = \Sigma N_i \qquad (83)$$

From $\bar{x}$, $s$, and $l$ we compute $k_o$, the observed difference between $U$ and the sample mean in units of the sample standard deviation

$$k_o = \frac{U - \bar{x}}{s} \qquad (84)$$

Then, given a confidence coefficient $\gamma$, a confidence interval which in a fraction $\gamma$ of samples will overlap the true fraction defective may be obtained from formulas (42) and (43) by substituting $k_o$ for $k$ and $(1 - \gamma)/2$ for $L_p$. The $p$ and $p'$ given by (42) and (43) are the upper and lower confidence limits $\bar{p}_\gamma$ and $\check{p}_\gamma$, respectively, for the fraction defective produced by the process. The larger the confidence coefficient, the broader the interval; and the smaller the confidence coefficient, the narrower the interval. In fact, if $\gamma$ is 1, the interval runs from 0 to 100 percent; and if $\gamma$ is 0, the interval includes only the point $p = \bar{p}$, where $\bar{p}$ is the probability that a normal deviate will exceed $k_o$.

If the acceptable quality level lies above the confidence interval, reduced inspection is permissible.* If it lies below the confidence interval, tightened inspection is necessary. If it lies within the confidence interval, normal inspection may be continued.

In order to avoid too frequent shifts in the degree of inspection, it is advisable that the criteria for resuming normal inspection when tightened or reduced inspection is in effect should differ from the criteria for remaining on normal inspection when it is already in effect. One way to achieve this is to use a larger confidence coefficient when deciding whether to continue normal inspection than when deciding whether to resume nor-

---

* For reduced inspection, it is customary to impose a further requirement to restrict the variability of the process average from lot to lot. This usually takes the form of requiring that each of the last twenty (or some other number) lots have been acceptable.

mal inspection. Reduced inspection, for example, will be installed as soon as the process average falls to a certain level, but once it has been installed normal inspection will not be resumed unless this level is exceeded by a certain margin. It is therefore possible, once either reduced or tightened inspection has been installed, to continue it under conditions that would not justify its introduction were normal inspection in effect.

**8.2.2. Illustrative Computation of Process Average.**—Suppose that the combined sample from several inspection lots contains 95 observations, that their sum is 2,872, that the sum of their squares is 89,175, and that values above 35 represent defects. Suppose a confidence coefficient $\gamma_c$ of 0.99 is used for judging whether to continue normal inspection and a confidence coefficient $\gamma_r$ of 0.95 for judging whether to resume normal inspection. Thus

$$N = 95 \qquad \Sigma x = 2{,}872 \qquad \Sigma x^2 = 89{,}175 \tag{85}$$

$$U = 35 \qquad \gamma_c = 0.99 \qquad \gamma_r = 0.95 \tag{86}$$

Then, from formulas (4) and (6)

$$\bar{x} = 30.2316 \qquad s = 5.000 \tag{87}$$

so

$$k_o = \frac{35 - 30.2316}{5} = 0.9537 \tag{88}$$

Entering Table 1.5 with $K_e = 0.9537$, we find (by linear interpolation) 17.01 percent as the point estimate $\bar{p}$, corresponding with a confidence coefficient of zero, of the process average percent defective.

To find the 99 percent confidence limits, we substitute $k = 0.9537$, $L_p = (1 - 0.99)/2 = 0.005$ into (42) and (43). The details of a similar computation are shown in Sec. 6.3.1. The 99 percent confidence limits are

$$\check{p}_{0.99} = 10.15 \text{ percent} \qquad \text{and} \qquad \hat{p}_{0.99} = 26.29 \text{ percent} \tag{89}$$

Similarly, taking $L_p = 0.025$, the 95 percent confidence limits are

$$\check{p}_{0.95} = 11.57 \text{ percent} \qquad \text{and} \qquad \hat{p}_{0.95} = 23.86 \text{ percent} \tag{90}$$

If normal inspection is in effect, it will be continued if the AQL for the product is between 10.15 and 26.29 percent defective; reduced inspection will be installed if the AQL exceeds 26.29 percent and tightened inspection will be installed if the AQL is below 10.15 percent. If tightened or reduced inspection is in effect, normal inspection will be resumed if the AQL is between 11.57 and 23.86 percent. If the difference between the intervals for continuation and for resumption is too small, that is, if too many shifts occur, it can be widened by increasing the difference between the two confidence coefficients used.

For routine operations in which the same AQL and the same value of $U$ are to be used repeatedly, it is desirable to prepare a table showing for various sample sizes the values of $k_o$ that result in continuing normal inspection if it is in use and those that result in resuming normal inspection if tightened or reduced inspection is in use. In the present illustration, for example, if the AQL is 15 percent, a value of $k_o$ computed from 95 observations leads under normal inspection to continuing normal inspection if it lies between $\check{k}_{0.99} = 0.7380$ and $\hat{k}_{0.99} = 1.4107$, or leads under reduced or tightened inspection to resuming normal inspection if it lies between $\check{k}_{0.95} = 0.8048$ and $\hat{k}_{0.95} = 1.3113$. These values are determined from formula (21), using both the plus and the minus signs to obtain upper and lower limits and taking $p_1 = $ AQL, $\alpha = (1 - \gamma)/2$. An example of the computation of (21) is given in Sec. 4.5.2.

### 8.3. Methods of Reducing or Tightening Inspection

Reduced inspection may be achieved by reducing $N$, perhaps as much as 40 or 50 percent, and adjusting $k$ so that the producer's risk will not be increased. This leaves the OC curve approximately unchanged for qualities equal to or better than the AQL but increases the risk that lots of worse quality will, if submitted, be accepted.

Tightened inspection may be achieved by increasing $k$ but leaving $N$ fixed. This moves the OC curve to the left, thus reducing the consumer's risk but increasing the producer's risk.

## 9. CHOICE OF QUALITY STANDARDS AND EVALUATION OF PROPOSED SAMPLING PLANS

### 9.1. General Discussion

The definition of a defective item, the choice of an AQL and an LTPD,* and the selection of producer's and consumer's risks usually require a compromise between what is technically or economically desirable in the use for which the product is intended and what is technically or economically feasible in its production. Although no fixed rules can be established for selecting $U$, $p_1$, $p_2$, $\alpha$, and $\beta$, a few guides are worth noting.

---

* LTPD, or lot tolerance percent defective, is defined sometimes in terms of the use for which the product is intended, as the worst quality for which there should be any appreciable probability of a lot's being accepted, and sometimes, in terms of the sampling plan, as that quality for which the probability of acceptance has a specified value (customarily 10 percent). Similarly, the AQL, or acceptable quality level, is defined sometimes as the worst quality that can be consistently accepted or as the best quality for which there should be any appreciable probability of a lot's being rejected, and sometimes, as that quality for which a certain sampling plan gives a specified probability of rejection (customarily 5 percent). We follow the definitions in terms of sampling plans, using the customary risks of 5 and 10 percent for the AQL and LTPD, respectively.

Some of these five quantities can be selected arbitrarily or by convention without restricting the range of available plans. For example, it is more or less conventional to take the producer's risk $\alpha$ as 5 percent and the consumer's risk $\beta$ as 10 percent. Then the AQL, or $p_1$, may be defined as that quality for which the probability of acceptance should be 95 percent and the LTPD, or $p_2$, as that quality for which the probability of acceptance should be 10 percent. Sometimes $\alpha$ and $\beta$ are set tentatively and values of $U$, $p_1$, and $p_2$ selected; then the values of $\alpha$ and $\beta$ may be reexamined in the light of the number of lots of qualities $p_1$ and $p_2$ likely to be submitted and the costs of rejecting those of quality $p_1$ or accepting those of quality $p_2$.

To select a sampling plan, two independent analyses should be made and their results merged. One study should consider the uses to which the product is to be put. What is required of each item, how rigid are these requirements, and what is the effect of a certain percentage of the items failing to meet the requirements? The other study should consider the production process. Within what limits do most of the items fall, how stable are these limits, what would be the costs of improving them, and what is the cost of having a lot rejected? In designing a sampling plan on the basis of these two studies, it is necessary to take into account the cost of inspection. What is the cost of drawing a sample, what is the cost of inspecting the items, what is the cost of analyzing the data, and how do all these vary with the size of the sample?

## 9.2. Analysis of Production Possibilities

It should be assumed that the producer will consider the OC curve as in effect the specification of what will be accepted, and that he will adjust his operations to an OC curve as best suits his own circumstances. For example, if the producer finds that a rejected lot entails considerable costs, he will consider that the given OC curve actually forces him to produce at a quality much better than the AQL, how much better depending upon how sharply the cost of production rises and the probability of a rejection declines as quality improves. If improvements in quality are also costly, a given OC curve may mean that the consumer will receive— and, of course, ultimately pay for—better quality than he needs. Under such circumstances the consumer will find it advantageous to increase the amount of inspection, for which also he ultimately pays; this will steepen the OC curve, that is, increase the rate at which the probability of a rejection declines as quality improves, and therefore reduce the degree to which the producer is forced to improve the quality and raise the cost.*

There is obviously a relation between $U$ on the one hand and $p_1$

---

* Some discussion of this point is given in J. Przyborowski and H. Wileński, "Testing Clover Seed," especially Sec. III(b), pp. 281–285.

and $p_2$ on the other. If $U$ is set very high, only very bad items are considered defective; only a small percentage of such items can be accepted, so $p_1$ should be small. If $U$ is set very low, items are considered defective when they deviate very little from normal; the effect of such items is less serious, so a larger percentage of them can be accepted.

If we regard $U$ as fixed, we may, by analyzing data on the quality characteristic that is to be inspected, estimate what proportion defective is being produced. Section 8.2 gives a method for finding confidence intervals $\check{p}_\gamma$ to $\hat{p}_\gamma$ which, in a proportion $\gamma$ of the samples from which such confidence intervals are computed, overlap the true fraction defective; or, if the confidence coefficient is reduced to zero, we may obtain a single-valued estimate $\bar{p}$. The height of the OC curve over the interval $\check{p}_\gamma$ to $\hat{p}$, then indicates approximately how much of the current production would be passed by a proposed sampling plan.

If, on the other hand, we regard $p_1$ as fixed, we may estimate the value of $U$ necessary to give a confidence $\gamma$ that the fraction defective will not exceed $p_1$. This value is

$$\hat{U}_\gamma = \bar{x} + \hat{k}_\gamma s \tag{91}$$

where $\hat{k}_\gamma$ is given by formula (21), using the plus sign and taking $p_1$ as the fixed fraction defective and $\alpha$ as $1 - \gamma$. Thus, in the example of Sec. 8.2.2, where $p_1 = 0.15$, $\bar{x} = 30.232$, and $s = 5.000$, $\hat{k}_{0.95} = 1.2632$ and $\hat{k}_{0.99} = 1.3695$; so we may be 95 percent confident that under the conditions prevailing when the sample was collected at least 85 percent of the items lie below $\hat{U}_{0.95} = 36.54$, or 99 percent confident that at least 85 percent lie below $\hat{U}_{0.99} = 37.08$. $\hat{U}$ is called a "tolerance limit."*

## 9.3. Analysis of Consumption Requirements

In analyzing the consumption, or design, requirements of a product, account should be taken of the variability of other products or methods with which it will be associated in use. A discussion of this is beyond the scope of this chapter,† but one especially simple example may be mentioned.

---

* See Chap. 2 for a discussion of tolerance limits. Chapter 2, however, is concerned with two-sided tolerance limits, that is, values of $k$ such that at least a given proportion $P$ of the population is included between the limits $\bar{x} - ks$ and $\bar{x} + ks$.

An excellent discussion of tolerance limits, which in fact is the original source from which subsequent theoretical work has developed, is given in Shewhart, *Statistical Method*, Chap. II, pp. 50–79, How Establish Limits of Variability?

† See Shewhart, *Economic Control of Quality*, Part V, pp. 247–272, Statistical Basis for Specification for Standard Quality, especially Chap. XVII, pp. 249–261, Design Limits on Variability, in which Sec. 5, pp. 252–256, considers Tolerances for Quality of Finished Product in Terms of Tolerances of Piece-parts, and Sec. 6, pp. 256–259, discusses The General Problem of Setting Tolerances on Controlled Product.

Suppose the items from a lot are to be assembled at random in sets of $n$, and a set of $n$ will be defective if the sum of the lengths of the $n$ pieces exceeds $nU'$, that is, if the mean length in the set exceeds $U'$. If we consider lot quality measured by the percent of means of $n$ that exceed $U'$ and select two points on the desired OC curve, values of $N$ and $k$, say $N'$ and $k'$, can be obtained from (10) and (11) or (12). If now inspection is carried out on the basis of individual items, rather than sets of $n$, the lot can be accepted if in a sample of $nN'$ items

$$\bar{x} + \frac{k'}{\sqrt{n}} s \leq U' \tag{92}$$

where $\bar{x}$ and $s$ are the mean and standard deviation computed from the $nN'$ items.

If in this example a piece were considered defective if it exceeded one-$n$th the allowable sum for $n$ pieces, that is, if it exceeded $U'$, the sampling plan would require a sample of $N'$ items and the acceptance requirement would be

$$\bar{x} + k's \leq U' \tag{93}$$

This would necessitate production at a higher quality than would (92). For example, if $n = 4$ and $k = 0.7645$, the indifference quality, that is, the quality having an equal chance of acceptance or rejection, is 35.12 percent defective for (92) and 22.23 percent for (93). If $n = 9$ and $k = 2.0077$ (the $k$ given in Table 1.2 for $p_1 = 0.015$, $p_2 = 0.03$), the indifference quality is 25.17 percent under (92) and 2.23 percent under (93).

As a digression, it may be observed that averages in sets of $n$ may be appropriate measures of quality when inspection is undertaken to determine whether full measure is delivered. Suppose tomato juice is packed in one-gallon cans* that the final purchasers buy one at a time. Then any purchaser whose can contains less than one gallon is cheated, whether or not some other purchaser receives an equivalent or even a greater excess. Even though the purchaser would not be cheated in the long run if his purchase were repeated frequently, an appreciable shortage in any one can may inconvenience or annoy him—in fact, may prevent his repeating the purchase. In such circumstances, the appropriate measure of lot quality is the percent of cans containing less than a stated amount, say one gallon, or perhaps 0.99 gallon if a 1 percent deviation is permissible. If, however, the tomato juice is to be sold to restaurants in cases of 12 cans, the appropriate measure of lot quality is the percent of groups of 12 that average less than one gallon per can. If the tomato juice is to be sold to Army camps in carloads, the percent of carloads

---

* The example is taken from Peach, *Industrial Statistics*, p. 138. Peach, however, treats the example differently.

averaging less than one gallon per can is the appropriate measure of lot quality; in such a case, however, one carload might well be an inspection lot, so that the lot mean would become the appropriate quality measure.

## 9.4. Normality Assumption

If within lots the distribution of items is not approximately normal, the sampling plan usually controls the percent defective as consistently as if the distribution were normal, but not at the level intended.

What the plan actually controls is the quantity defined by equation (1), namely,

$$K = \frac{U - \mu}{\sigma} \qquad (94)$$

Thus a lot is rejected if the sample value of $K$, say $k_o$, is too small and accepted if it is sufficiently large. The OC curve might be plotted with the horizontal axis representing the true value of $K$ in submitted lots; such an OC curve would be essentially correct regardless of the normality of the distribution within lots. The assumption of normality becomes important when translating values of $K$ into values of $p$, the fraction defective.

If the distribution is not normal but its *form* is stable from lot to lot and is such that the percent defective is determined by $K$, then it is possible to substitute for Table 1.5 another table more appropriate for translating $K$ into percent defective. In any event, it may be assumed that the fraction defective corresponding with any $K$ does not exceed $(1/K)^2$; and if the distribution is unimodal and has its mode near its mean it may be assumed that for any $K$ the fraction defective does not exceed $(2/3K)^2$.[*]

The effects of non-normality are excellently stated by Jennett and Welch:[†]

The above theory is developed on the assumption that the parent population sampled is Gaussian. Even if it is not, the reasoning relating the distribution of $k_o$ with $K$ will not be invalidated unless the samples are very small, because non-normality will not have much influence on $\bar{x}$ and $s$. The chief error lies in connecting $K$ with $p$ by using the normal tables. This can be obviated if some other constant distributional form can be assumed for which the relation between the standardized deviate $K$ and $p$ is known. If we could say that the distribution would maintain the same shape . . . , although with possible changes in the mean and standard deviation, the above theory would need little modification.

---

[*] These quantities are based on the Tchebycheff and Camp-Meidell inequalities, respectively (see Shewhart, *Economic Control of Quality*, pp. 176–177); in the present context they can be halved if symmetry can be assumed, and reduced further if positive skewness is assumed.

[†] "The Control of Proportion Defective." The passage quoted is from Sec. 3, p. 88, Effect of Non-normality. The notation in the passage as quoted has been altered to agree with that used in this chapter.

It would be misleading, however, to lay too much emphasis on the necessity for making the connection between $K$ and $p$ correctly. Usually it is not possible to say with absolute certainty that a particular value of $p$ is satisfactory and another, differing little from it, is unsatisfactory. Indeed, often it may not be clear that $p$ has any intrinsic importance at all, but it may still be useful for purposes of control. For the same reason $K$ may be regarded as a useful feature of the distribution to aim at controlling, without inquiring too closely into the exact meaning of $K$. Non-normality is therefore not such a serious matter as it may appear. However, no final word can be said about this without reference to the context in which the method is to be used.

## 10. SIMULTANEOUS USE OF SEVERAL SAMPLING PLANS

### 10.1. Inspection of Several Quality Characteristics

Often the quality of an item depends upon several variables. There may be measurements of physical dimensions, for example, those which determine whether the item will fit properly in a final assembly, or whether full measure has been delivered; measurements of the properties of the item, for example, hardness, chemical composition, tensile strength, or modulus of elasticity; measurements of performance, for example, duration of useful life or rate of output; and so forth. The lot quality then is measured by $n$ fractions defective, say $p_1, p_2, \ldots, p_n$.

Suppose that $n$ criteria of the form

$$\bar{x} + ks \leq U \tag{95}$$

are applied to a single sample of $N$. The probability that the lot will pass all $n$ tests cannot be read from the OC curve for any one test. For a given set of values $p_1, p_2, \ldots, p_n$, the probability that a lot will be accepted cannot be more than the smallest among the $n$ values of $L_p$; it will be this large if, and only if, the $n$ tests are perfectly correlated, so that a lot passing any one test is certain to pass all. At the other extreme, the probability of acceptance may be zero; it will be this small if a lot that passes one test is certain to fail some other, or if $L_p$ is zero for any of the $n$ values of $p$.*

---

* It might be thought that these two extreme cases would never be encountered in practice. Specification writing and acceptance test designing being what they are, this is not the case. The author cherishes the memory of one particular specification for a cylindrical tube. It specified length, inside diameter, outside diameter, concentricity, weight, density, absence of voids, and all the details of chemical composition, physical properties, and manufacturing processes that determine density. Such a specification must be, at least in some degree, either redundant or inconsistent; this particular one perhaps had been redundant originally, but through an encrustation of changes made in one detail at a time it had become inconsistent. It would be naïve to assume, however, that because no material could meet the acceptance criteria none was accepted. On the contrary, all of it was accepted. This was achieved partly by the buyer's reasoning that the product was needed, that the acceptance tests failed to pass usable material, and that therefore material failing to pass the tests was usable;

Between these two extremes of perfect positive correlation and perfect negative correlation among the tests is the case in which the tests are independent.   In this case, the probability of accepting a lot of quality $p_1$, $p_2$, . . . , $p_n$ is given by the product of the individual probabilities of acceptance:

$$L_{p_1, p_2, \ldots, p_n} = L_{p_1} \cdot L_{p_2} \cdot \cdots \cdot L_{p_n} \tag{96}$$

Since each $L_{p_i}$ is 1 or less, this product is less than the least $L_{p_i}$ among the set of $n$.   If $n = 3$ and each $L_{p_i} = 0.95$, the product is $(0.95)^3$ or $0.857$; if each $L_{p_i} = 0.99$, the product is $(0.99)^3$ or $0.970$.   For $n = 10$ the corresponding products are $0.599$ and $0.904$.   (When $L_{p_i}$ is near 1 and $n$ is not too large, the probability of rejecting is about $n$ times as large for $n$ independent tests as for one.)

Sometimes the situation is that for good quality the various tests are independent but that for poor quality the tests are positively correlated. Thus, if for each of five characteristics a lot has fractions defective that give it a 95 percent chance of acceptance on the individual tests, its chance of passing all five tests is about $(0.95)^5$ or $0.774$.   But if its chance of acceptance on each individual test is $0.10$, its chance of acceptance on all five may be nearly $0.10$.   In such situations, not only is the OC curve for the battery of tests lower than that for any one test, but it gives less sharp discrimination between good and bad lots.   On the other hand, if the tests are independent or positively correlated for good quality but negatively correlated for bad quality, the discrimination is sharpened.

Another aspect of the simultaneous use of several criteria like (95) is that this treats the various tests on an attribute basis, so to speak.   If a lot just barely fails to pass one test but appears to be of exceptionally high quality on all other tests, it is rejected; whereas a second lot that just barely passes each test is accepted.   It would be preferable to combine all the measurements on one item into a single score for that item and judge the lot by the scores of the sample items. *

All these uncertainties about the simultaneous operation of several criteria make the selection, operation, and interpretation of the plan much less clear-cut and straightforward than when only one criterion is con-

and partly by the seller's disclaiming responsibility for the quality of the product, reasoning that he could take responsibility either for the quality of the product or for following detailed instructions for making it, but not for the consistency of those instructions with the desired quality.

* Chapter 3 deals with this problem.   The methods presented there are new and have not yet been used extensively or developed to the point where they are comparable with those of this chapter in simplicity of selecting plans and constructing OC curves, or in ease of routine operation and interpretation.   They will nevertheless well repay careful study, especially when complicated and expensive products or experiments are concerned.

cerned. They underscore the necessity, if variables are to be used, of a careful preliminary analysis of a rather large body of data on the product to which the plans are to be applied.

## 10.2. Upper and Lower Bounds on a Single Quality Characteristic

### 10.2.1. General Considerations.—The OC curve for two simultaneous criteria of the form (96) can be determined mathematically in the case where the two criteria simply impose upper and lower limits on a single characteristic; that is, the case where an item is defective if it is either above some upper limit $U$ or below some minimum $M$. Some mathematical investigations of this have been made,* but no completely satisfactory solution has yet been obtained. The indications are, however, that for practical purposes it will ordinarily suffice to use two criteria of the form (96) determined independently for the upper and lower limits. Two points about this procedure, however, should be recognized.

The first point is that the combinations of $\bar{x}$ and $s$ that lead to acceptance are not quite the same when two one-sided criteria are used as they would be if a single two-sided test were constructed. The combinations that lead to acceptance with two one-sided tests (having the same values of $N$ and $k$) are included in an area shaped like the triangle of Fig. 1.8, whereas a single two-sided test would result in rounding off the top corner of this triangle, somewhat as in the curve of Fig. 1.8. If most of the production should fall in the small area not common to both regions, this difference would be of importance. This would be a situation in which the mean behaved excellently, consistently lying midway between $M$ and $U$, but the standard deviation was continually near the maximum value permitting acceptance with the best value of the mean. If such a situation were to prevail consistently, acceptance inspection would probably be based on the standard deviation alone.

The second point to be kept in mind in using two one-sided tests for a given characteristic is the effect on the OC curve. The true OC curve for a two-sided test cannot be expressed in terms of the lot fraction defective alone; it depends upon the lot mean and standard deviation. A sample that fails to meet one criterion because of a large standard deviation is likely to fail the other criterion also; but a sample that fails one criterion because of an extreme mean is likely to pass the other criterion.

Thus, although the use of two one-sided criteria to control both upward and downward variation often works out satisfactorily in practice, it should be recognized as an improvisation for a problem that is in essential respects different from the one-sided problem and is less amenable to simple methods and interpretations.

* The mathematical work was done in the Statistical Research Group by A. Wald and J. Wolfowitz but has not been published.

**10.2.2. Use of a Pair of One-sided Criteria.**—Suppose two tests are constructed so that with one the probability is $\beta$ that a lot with fraction $p$ ($p < 0.5$) above $U$ will be accepted and with the other the probability is $\beta$ that a lot with fraction $p$ below $M$ will be accepted. If the standard deviation of the measurements is small, the probability is practically $\beta$ that a lot with fraction $p$ outside the limits $M$ and $U$ will be accepted; but if the standard deviation is large, the probability is practically $\beta$ that a lot with fraction $2p$ outside the limits $M$ and $U$ will be accepted. If $p/2$ is substituted for $p$, the test will give the intended probability of acceptance for high values of the standard deviation but not for low values of the standard deviation.

If the standard deviation is expected always to be low—it need not be so low when the relevant values of $p$ are small as when they are large—or if lots with high standard deviations are sufficiently infrequent as to have only negligible effects on the operation of the tests, a pair of one-sided tests may be used successfully. The two acceptance criteria, both of which must be satisfied before a lot is accepted, are then

$$\bar{x} + ks \leq U \qquad \text{and} \qquad \bar{x} - ks \geq M \tag{97}$$

It may be noted that if $\bar{x} > (M + U)/2$, the second inequality can be ignored, for it will surely be met if the first is; similarly, if $\bar{x} < (M + U)/2$, the first inequality can be ignored.

As in the case of a single criterion, we may note that if

$$X > NU \qquad \text{or} \qquad X < NM \tag{98}$$

the lot may be rejected without regard to $Y$, where $X$ and $Y$ are the sum of the observations and the sum of the squares of the observations, respectively, as defined by formulas (51) and (52).

Second, we may note that if

$$Y \leq \frac{U^2(N-1)}{k^2 + 1} \qquad \text{and} \qquad X \geq N\frac{M+U}{2} \tag{99}$$

the lot may be accepted.

Third, if neither (98) nor (99) is satisfied, the lot is accepted if

$$X \geq N\frac{M+U}{2} \qquad \text{and} \qquad Y \leq aX^2 - b_U X + c_U \tag{100}$$

or

$$X \leq N\frac{M+U}{2} \qquad \text{and} \qquad Y \leq aX^2 - b_M X + c_M \tag{101}$$

where

$$a = \frac{Nk^2 + N - 1}{N^2k^2} \tag{102}$$

$$b_U = \frac{2U(N-1)}{Nk^2} \qquad b_M = \frac{2M(N-1)}{Nk^2} \tag{103}$$

$$c_U = \frac{U^2(N-1)}{k^2} \qquad c_M = \frac{M^2(N-1)}{k^2} \tag{104}$$

If neither (100) nor (101) is met, the lot is rejected.

For adjusted observations, as defined by (71), the lot may be accepted without regard to $X$ if

$$Y' \le \frac{(U-A)^2(N-1)}{k^2+1} \qquad \text{when} \qquad A \ge \frac{M+U}{2} \tag{105}$$

or

$$Y' \le \frac{(A-M)^2(N-1)}{k^2+1} \qquad \text{when} \qquad A \le \frac{M+U}{2} \tag{106}$$

In place of (100) and (101) we have

$$X \ge N\,\frac{M+U}{2} \qquad \text{and} \qquad Y' \le aX^2 - (b_U + 2A)X + (c_U + NA^2) \tag{107}$$

$$X \le N\,\frac{M+U}{2} \qquad \text{and} \qquad Y' \le aX^2 - (b_M + 2A)X + (c_M + NA^2) \tag{108}$$

The most efficient form of (105) and (106), either for graphical purposes or for permitting immediate acceptance on the basis of $Y'$ alone, is obtained by setting

$$A = \frac{M+U}{2} \tag{109}$$

In this case inequalities (105) and (106) are identical.

For either adjusted or unadjusted observations graphs may be constructed as described in Sec. 7.3. The maximum value of $Y$ for acceptance for each value of $X$ will be obtained from formula (100) or (107) for values of $X$ greater than $N(M+U)/2$ and from formula (101) or (108) for values of $X$ less than $N(M+U)/2$. In the case of adjusted observations, if $A = (M+U)/2$ the figures will be symmetrical about the line $X = NA$. The upper boundaries of the error region may be found by formula (70) or (74) of Sec. 7.3.

**10.2.3. Two-sided Criterion.**—A two-sided criterion is most easily applied if expressed in terms of a chart similar to Fig. 1.8, in which values of $\bar{x}$ are represented horizontally and values of $s$ vertically. For each sample, a point is plotted corresponding to the observed $\bar{x}$ and $s$. If this point lies within the region bounded by the acceptance curve and the horizontal axis, the lot is accepted; otherwise, it is rejected.

The first step in finding the boundary of the acceptance region is to compute $k$ and $N$ from formulas (10) and (11). The values of $p_1$ and $p_2$ now represent fractions outside the interval $M$ to $U$, instead of fractions above $U$.

The second step is to find the indifference quality $p_0$, which is simply the probability that a normal deviate will exceed $k$. It may be determined from Table 1.5.

The third step is to divide $p_0$ arbitrarily into two parts, $p_0'$ and $p_0''$ ($p_0' \leq p_0''$) such that

$$p_0' + p_0'' = p_0 \tag{110}$$

Each pair $p_0'$, $p_0''$ leads to two points on the acceptance curve, so enough different pairs should be used to permit plotting the curve.
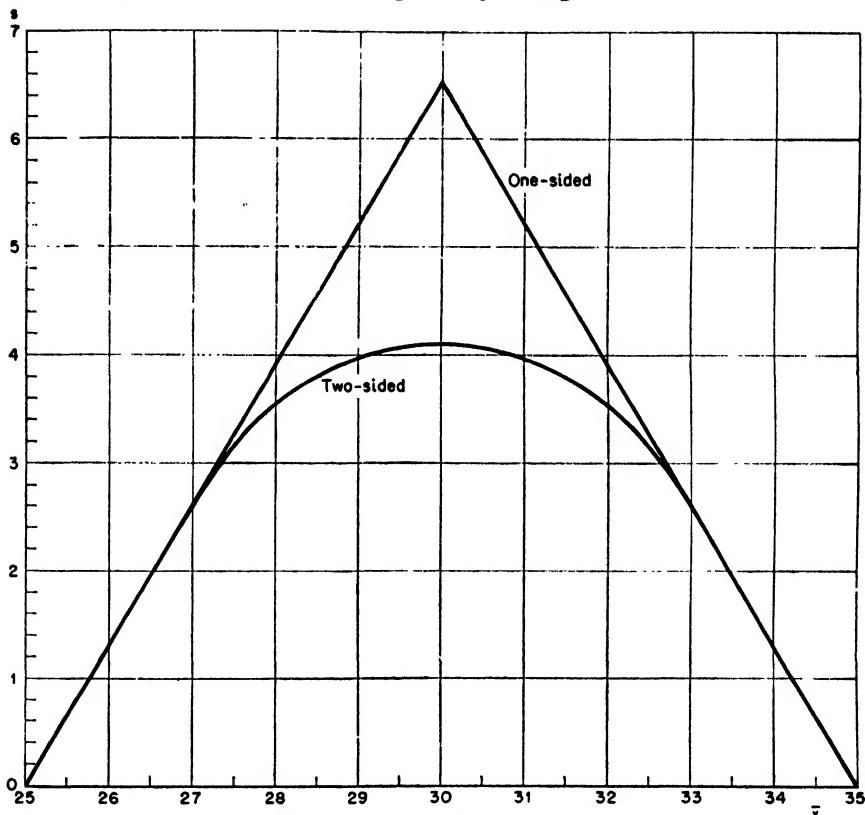


FIG. 1.8.—Acceptance region for a two-sided test defined by $p_1 = 0.15$, $p_2 = 0.30$, $\alpha = 0.01$, $\beta = 0.02$, $U = 35$, and $M = 25$; also corresponding region for a pair of one-sided tests. [*Data from Table 1.7, columns (9), (10) and (11).*]

The fourth step is to find $K_{p_0'}$ and $K_{p_0''}$, normal deviates exceeded with probabilities $p_0'$ and $p_0''$, respectively. These may be obtained from Table 1.1.

The fifth step is to substitute each pair of values $K_{p_0'}$, $K_{p_0''}$ into the pair of equations

$$\bar{x} = \frac{UK_{p_0'} + MK_{p_0''}}{K_{p_0'} + K_{p_0''}} \qquad s = \frac{U - M}{K_{p_0'} + K_{p_0''}} \tag{111}$$

TABLE 1.7

COMPUTATION OF BOUNDARY OF ACCEPTANCE REGION FOR A TWO-SIDED TEST DEFINED BY $p_1 = 0.15$, $p_2 = 0.30$, $\alpha = 0.01$, $\beta = 0.02$,
$U = 35$, $M = 25$ ($k = 0.7645$; $p_0 = 0.2223$)

| $p_0'$ | $p_0'$ | $K_{p_0}'$ | $K_{p_0}''$ | (3) + (4) | (3) − (4) | (6) ÷ (5) | 2 ÷ (5) | $\bar{x} = \dfrac{U+M}{2} + \dfrac{U-M}{2} \cdot$ (7) | $\bar{x}' = U + M -$ (9) | $s = \dfrac{U-M}{2} \cdot$ (8) |
|---|---|---|---|---|---|---|---|---|---|---|
| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| 0.11115 | 0.11115 | 1.2204 | 1.2204 | 2.4408 | 0.0000 | 0.0000 | 0.8194 | 30.000 | 30.000 | 4.097 |
| 0.0912 | 0.1311 | 1.3334 | 1.1212 | 2.4546 | 0.2122 | 0.0864 | 0.8148 | 30.432 | 29.568 | 4.074 |
| 0.0712 | 0.1511 | 1.4669 | 1.0317 | 2.4986 | 0.4352 | 0.1742 | 0.8004 | 30.871 | 29.129 | 4.002 |
| 0.0512 | 0.1711 | 1.6333 | 0.9498 | 2.5831 | 0.6835 | 0.2646 | 0.7743 | 31.323 | 28.677 | 3.872 |
| 0.0312 | 0.1911 | 1.8634 | 0.8738 | 2.7372 | 0.9896 | 0.3615 | 0.7307 | 31.808 | 28.192 | 3.654 |
| 0.0112 | 0.2111 | 2.2835 | 0.8026 | 3.0861 | 1.4809 | 0.4799 | 0.6481 | 32.400 | 27.600 | 3.240 |
| 0.0001 | 0.2222 | 3.7190 | 0.7648 | 4.4838 | 2.9542 | 0.6589 | 0.4461 | 33.294 | 26.706 | 2.230 |

NOTE: The data of columns (9), (10) and (11) are charted in Fig. 1.8, p. 55.

or into the equivalent pair

$$\bar{x} = \frac{U + M}{2} + \frac{U - M}{2} \cdot \frac{K_{p_0'} - K_{p_0''}}{K_{p_0'} + K_{p_0''}} \qquad s = \frac{U - M}{2} \cdot \frac{2}{K_{p_0'} + K_{p_0''}} \qquad (112)$$

Each pair of values $K_{p_0'}$, $K_{p_0''}$ when substituted into (111) or (112) leads to a combination $\bar{x}, s$ lying on the boundary of the acceptance region; another point on the boundary is given by the combination $\bar{x}', s$ where

$$\bar{x}' = U + M - \bar{x} \qquad (113)$$

Table 1.7 shows the computations for the two-sided acceptance region defined by $p_1 = 0.15$, $p_2 = 0.30$, $\alpha = 0.0\overline{1}$, and $\beta = 0.02$, where defects are represented by values below 25 or above 35. The values $N = 95$ and $k = 0.7645$ are taken from Sec. 4.3. The resulting region is plotted in Fig. 1.8. The figure also shows the acceptance region given by using the pair of one-sided criteria

$$\bar{x} - 0.7645s \geq 25 \qquad \bar{x} + 0.7645s \leq 35 \qquad (114)$$

It will be observed that the acceptance region for the two-sided test lies entirely within that for the pair of one-sided tests. As $\bar{x}$ approaches $U$ or $M$, the two boundaries nearly coincide. The difference between the regions is fairly sizable in this case, which involves rather large fractions defective, but for smaller fractions defective the two regions more nearly coincide. For example, keeping $M = 25$ and $U = 35$, but taking $p_1 = 0.001$, $p_2 = 0.0015$, $\alpha = 0.05$, $\beta = 0.10$ (for which Table 1.2 shows $k = 3.0214$) we find that for $\bar{x} = 30$ the maximum acceptable value of $s$ is given by the pair of one-sided tests as 1.655 and by the two-sided test as 1.550; and the difference is less for any other value of $\bar{x}$. The effect of the difference between a pair of one-sided tests and a two-sided test should be judged by the frequency of lots leading to sample points that would signify acceptance by one test and rejection by the other. One way in which the two-sided test presented here is useful is in judging whether two one-sided tests can be used satisfactorily.

If a two-sided test is used, criteria for reducing or tightening inspection may also be expressed graphically. Upper and lower limits of $k_o$ should be computed as described in the last paragraph of Sec. 8.2.2 and used as values of $k$ from which to compute two boundary curves according to formulas (111) or (112). If the value of $k_o$ found in computing the PA lies between the two curves, normal inspection is continued; if it falls below the lower curve, tightened inspection is installed; and if it falls above the upper curve reduced inspection is installed.

## 11. TECHNICAL NOTES

### 11.1. Basic Formulas

**11.1.1. Assumptions.**—Sampling plans for controlling percent defective on the basis of variables involve drawing a sample of $N$ at random

from a lot and computing

$$\bar{x} + ks \tag{115}$$

where .

$$\bar{x} = \frac{\Sigma x}{N} \tag{116}$$

$$s^2 = \frac{N \Sigma x^2 - (\Sigma x)^2}{N(N-1)} \tag{117}$$

$x$ represents measurements of the variable being inspected, and $k$ is a constant.

If $x$ is normally distributed, $\bar{x}$ will be normally distributed with mean $\mu$ and standard deviation $\sigma/\sqrt{N}$, where $\mu$ and $\sigma$ are the lot mean and standard deviation of $x$. Also, unless $N$ is extremely small, say, less than 5, $s$ will also be nearly normally distributed, its mean and standard deviation being $a\sigma$ and $b\sigma/\sqrt{2(N-1)}$, respectively, where

$$a = \frac{\sqrt{\dfrac{2}{N-1}}\ \Gamma\left(\dfrac{N}{2}\right)}{\Gamma\left(\dfrac{N-1}{2}\right)} \tag{118}$$

$$b = \sqrt{2(N-1)(1-a^2)} \tag{119}$$

The factors $a$ and $b$ are slightly less than 1 — when $N = 5$, for example, $a = 0.940$ and $b = 0.965$ — and they increase toward 1 as $N$ increases.* They are therefore treated as 1 throughout this chapter; that is, the standard deviation is regarded as normally distributed with mean $\sigma$ and standard deviation $\sigma/\sqrt{2(N-1)}$.

Since a linear combination of independently distributed variables tends to be normally distributed even when the individual variables depart from normality, $\bar{x} + ks$ will be more nearly normally distributed than $s$; in fact, $\bar{x} + ks$ will tend toward normality even if the distribution of $x$ departs from normality. The mean of $\bar{x} + ks$ is $\mu + k\sigma$ and the standard deviation is

$$\sigma \sqrt{\frac{1}{N} + \frac{k^2}{2(N-1)}} \tag{120}$$

---

* A table of $a$ and $b$ is given in Jennett and Welch, "Control of Proportion Defective," p. 85. This table gives $a$ and $b$ to 3 decimal places for values of $N$ from 5 to 15 by steps of 1, from 15 to 30 by steps of 5, and for 40, 50, 75, and 100. A table of $c_1 = a \sqrt{(N-1)/N}$ to 5 decimal places for values of $N$ from 3 to 25 by steps of 1 and from 25 to 100 by steps of 5 is given by Shewhart, *Economic Control of Quality*, p. 185. Values of $c_1 = a \sqrt{(N-1)/N}$ are given to 4 decimal places for values of $N$ from 2 to 25 by steps of 1 in American Society for Testing Materials *Manual*, p. 50 of Supplement B; the same table is given in American Standards Association, *Control Chart Method*, p. 39.

**11.1.2. Direct Derivation.**—The probability that in a sample $\bar{x} + ks$ will be less than $U$ then depends on the difference between $U$ and $\mu + k\sigma$, measured in units of the standard deviation of $\bar{x} + ks$; that is, the probability that $\bar{x} + ks$ will be less than $U$ is the probability that a unit normal deviate will be less than

$$\frac{U - (\mu + k\sigma)}{\sigma\sqrt{\dfrac{1}{N} + \dfrac{k^2}{2(N-1)}}} \tag{121}$$

The probability will be $\epsilon$ that the test will result in acceptance if

$$\frac{U - (\mu + k\sigma)}{\sigma\sqrt{\dfrac{1}{N} + \dfrac{k^2}{2(N-1)}}} = -K_\epsilon = K_{1-\epsilon} \tag{122}$$

where $K_\epsilon$ is defined by

$$\frac{1}{\sqrt{2\pi}} \int_{K_\epsilon}^{\infty} c^{-\frac{t^2}{2}} dt = \epsilon \tag{123}$$

We may write (122)

$$\frac{U - \mu}{\sigma} - k = -K_\epsilon \sqrt{\frac{1}{N} + \frac{k^2}{2(N-1)}} \tag{124}$$

$(U - \mu)/\sigma$ is a unit normal deviate that will be exceeded with some probability $p$, where $p$ is the proportion defective in the lot, so we may write

$$K_p = \frac{U - \mu}{\sigma} \tag{125}$$

Substituting this into (124) and rearranging, we find

$$k = K_p + K_\epsilon \sqrt{\frac{1}{N} + \frac{k^2}{2(N-1)}} \tag{126}$$

If both sides of (126) are decreased by $K_p$ and then squared, a rearrangement of terms gives

$$\left(1 - \frac{K_\epsilon^2}{2(N-1)}\right) k^2 - 2K_p k + \left(K_p^2 - \frac{K_\epsilon^2}{N}\right) = 0 \tag{127}$$

Denoting the coefficient of $k^2$ by $a$ and the term independent of $k$ by $b$, $k$ may be expressed in terms of $N$, $K_p$, and $K_\epsilon$ as

$$k = \frac{K_p + \sqrt{K_p^2 - ab}}{a} \tag{128}$$

If $K_p + k$ is subtracted from both sides of (126) and the equation

multiplied by $-1$, we have an equation expressing the fraction defective, in terms of a normal deviate, as a function of $N$, $k$, and $\epsilon$; that is, the abscissa of the OC curve is expressed as a function of the ordinate of the OC curve and of the parameters defining the sampling plan.

Suppose we desire to choose $N$ and $k$ so that the probability will be $1 - \alpha$ of accepting a lot of quality $p_1$ and $\beta$ of accepting a lot of quality $p_2$, where $\alpha + \beta < 1$ and $p_2 > p_1$. Substituting $p_1$ and $1 - \alpha$ into (126) gives one equation in $N$ and $k$, and substituting $p_2$ and $\beta$ gives another

$$k = K_1 + K_{1-\alpha} \sqrt{\frac{1}{N} + \frac{k^2}{2(N-1)}} \tag{129}$$

$$k = K_2 + K_\beta \sqrt{\frac{1}{N} + \frac{k^2}{2(N-1)}} \tag{130}$$

where $K_1$ and $K_2$ have been written instead of $K_{p_1}$ and $K_{p_2}$. After subtracting $K_1$ from (129) and $K_2$ from (130) and dividing (129) by (130), we have

$$\frac{k - K_1}{k - K_2} = \frac{K_{1-\alpha}}{K_\beta} \tag{131}$$

Solving this for $k$ and replacing $K_{1-\alpha}$ by $-K_\alpha$ gives

$$k = \frac{K_\alpha K_2 + K_\beta K_1}{K_\alpha + K_\beta} \tag{132}$$

$N$ may then be determined by substituting $k$ into (129) or (130) and solving. The solution will be considerably simpler and there will be little if any loss in accuracy if in the denominator of the second term under the radical $N - 1$ is replaced by $N$. With this approximation (129) becomes

$$k = K_1 + \frac{K_{1-\alpha}}{\sqrt{N}} \sqrt{1 + \frac{k^2}{2}} \tag{133}$$

Subtracting $K_1$ from both sides, multiplying by $\sqrt{N}/(K_1 - k)$, squaring, and replacing $K_{1-\alpha}$ by $-K_\alpha$ yields

$$N = \frac{K_\alpha^2 \left(1 + \frac{k^2}{2}\right)}{(K_1 - k)^2} \tag{134}$$

Similarly, (130) leads to

$$N = \frac{K_\beta^2 \left(1 + \frac{k^2}{2}\right)}{(k - K_2)^2} \tag{135}$$

By adding (134) and (135) and replacing the $k$ of the denominators by

(132), $N$ may be expressed as in (12); and by substituting for $k$ in the numerators as well as the denominators, $N$ may be expressed directly in terms of $K_1$, $K_2$, $K_\alpha$, and $K_\beta$, as in (11).

If instead of replacing $N - 1$ by $N$ in (129) and (130) we replace $N$ by $N - 1$, the right-hand sides of (134) and (135) are not affected; but the left-hand sides become $N - 1$. This approximation therefore leads to a value of $N$ one unit larger than that given by (134) and (135). Since $N$ must be an integer, the effects of rounding are of the same order of magnitude as those of replacing $N - 1$ by $N$. It is because of this approximation that values of $N$ given by (134) and (135) that are not integers should be rounded up instead of to the nearest integer. If the approximation is not made, we find

$$N = \frac{b + \sqrt{b^2 - 16a}}{4a} \tag{136}$$

where

$$a = \left(\frac{K_1 - K_2}{K_\alpha + K_\beta}\right)^2 \tag{137}$$

$$b = k^2 + 2(a + 1) \tag{138}$$

**11.1.3. Derivation from General Formulas for Sample Size and Critical Level.**—As an alternative derivation of formulas (132) and (134), we may start from the fact that $\bar{x} + ks$ is normally distributed about a certain mean in good lots and about another mean in bad lots and employ the ordinary procedures for testing the mean of a normally distributed variable. Let $z$ represent a normally distributed variable whose mean is $Z$ and whose standard deviation is a function of the sample size of the form

$$\sigma(N) = \sigma \cdot F(N) \tag{139}$$

where $F(N)$ is any function of $N$ that is independent of $\sigma$. Suppose a test is required such that the probability is $\alpha$ of rejecting a lot in which the mean is $Z_1$ and $\beta$ of accepting a lot in which the mean is $Z_2$, where $Z_2 > Z_1$. The test is made by finding $z$ in a sample of $N$ and accepting the lot if $z < A$ or rejecting it if $z > A$, where $A$ is a constant. Then if a lot has a mean of $Z_1$, the probability $L_1$ of acceptance is

$$L_1 = \frac{1}{\sigma(N)\sqrt{2\pi}} \int_{-\infty}^{A} e^{-\frac{(z - Z_1)^2}{2\sigma^2(N)}} dz \tag{140}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{A - Z_1}{\sigma(N)}} e^{-\frac{t^2}{2}} dt$$

where

$$t = \frac{z - Z_1}{\sigma(N)} \tag{141}$$

If the lot mean is $Z_2$ the probability $L_2$ of acceptance is

$$L_2 = \frac{1}{\sigma(N)\sqrt{2\pi}} \int_{-\infty}^{A} e^{-\frac{(z-Z_2)^2}{2\sigma^2(N)}} dz \tag{142}$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{A-Z_2}{\sigma(N)}} e^{-\frac{t^2}{2}} dt$$

where $t$ is defined by (141) but in terms of $Z_2$ instead of $Z_1$. The appropriate test is then defined by values of $A$ and $N$ such that

$$L_1 = 1 - \alpha \tag{143}$$
$$L_2 = \beta \tag{144}$$

that is,

$$\frac{A - Z_1}{\sigma(N)} = K_\alpha \tag{145}$$

$$\frac{A - Z_2}{\sigma(N)} = -K_\beta \tag{146}$$

On dividing (145) by (146) and solving for $A$, we find

$$A = \frac{K_\alpha Z_2 + K_\beta Z_1}{K_\alpha + K_\beta} \tag{147}$$

On multiplying both (145) and (146) by $\sigma(N)$, subtracting (146) from (145), replacing $\sigma(N)$ by $\sigma \cdot F(N)$, and solving for $F(N)$, we find

$$F(N) = \frac{Z_2 - Z_1}{\sigma(K_\alpha + K_\beta)} \tag{148}$$

In the ordinary case of testing the mean of a normally distributed variable, $F(N)$ is simply $1/\sqrt{N}$, and this leads to $N = 1/a$, where $a$ is defined by (137).

In applying formulas (147) and (148) to the case of percent defective, we have

$$z = \bar{x} + ks \tag{149}$$
$$Z = \mu + k\sigma \tag{150}$$

$$F(N) = \sqrt{\frac{1}{N} + \frac{k^2}{2(N-1)}} \tag{151}$$

A lot is considered good if

$$\frac{U - \mu}{\sigma} \geq K_1 \tag{152}$$

and bad if

$$\frac{U - \mu}{\sigma} \leq K_2 \tag{153}$$

where $K_1 > K_2$. From (152), using the equality sign, we find

$$\mu = U - K_1\sigma \tag{154}$$

Substituting this into (150) gives

$$Z_1 = U + (k - K_1)\sigma \qquad (155)$$

Similarly,

$$Z_2 = U + (k - K_2)\sigma \qquad (156)$$

Substituting from (151), (155), and (156) into (148), we have

$$\sqrt{\frac{1}{N} + \frac{k^2}{2(N-1)}} = \frac{K_1 - K_2}{K_\alpha + K_\beta} \qquad (157)$$

for which (136) is the solution. Substituting from (155) and (156) into (147) gives

$$A = \frac{K_\alpha[U + (k - K_2)\sigma] + K_\beta[U + (k - K_1)\sigma]}{K_\alpha + K_\beta} \qquad (158)$$

$$= \frac{K_\alpha(U + k\sigma) + K_\beta(U + k\sigma)}{K_\alpha + K_\beta} - \frac{K_\alpha K_2 + K_\beta K_1}{K_\alpha + K_\beta}\sigma$$

$$A = U + k\sigma - \frac{K_\alpha K_2 + K_\beta K_1}{K_\alpha + K_\beta}\sigma \qquad (159)$$

Any combination of values of $k$ and $A$ that satisfy this equation will be satisfactory values for the multiple of the standard deviation and the rejection level of $\bar{x} + ks$, but most pairs will be unusable because calculation of the rejection level $A$ from (159) will require knowledge of $\sigma$. If, however, we set

$$k = \frac{K_\alpha K_2 + K_\beta K_1}{K_\alpha + K_\beta} \qquad (160)$$

this problem is solved by the disappearance of $\sigma$, and the rejection level becomes simply

$$A = U \qquad (161)$$

The value of $k$ from (160) may then be substituted into (157) to determine $N$. If we replace $N - 1$ by $N$, the value of $N$ given by (157) is the same as that given by formulas (134) and (135), otherwise it is the same as that given by (136).

**11.1.4. Two-sided Criterion.**—The basis for the two-sided test presented in Sec. 10.2.3 is the fact that, because $\bar{x} + ks$ can be assumed symmetrically distributed about $\mu + k\sigma$, the boundary between acceptable and unacceptable combinations of $\bar{x}$ and $s$ corresponds with the boundary between combinations of $\mu$ and $\sigma$ for which a lot has more or less than a 50 percent chance of acceptance. Thus, in Fig. 1.8 if the $\bar{x}$ and $s$ axes are interpreted as $\mu$ and $\sigma$ axes, any lot represented by a point on the boundary line has an equal chance of acceptance or rejection, that is, corresponds with the indifference percent defective. For the two-sided test it is clear that the indifference quality is the same as for a one-sided test; for when $\sigma$ is very small and $\mu$ is near $U$ or $M$, the two-

sided and one-sided tests become virtually equivalent. Thus, the procedure for finding the boundary of the two-sided acceptance region begins by finding the proper indifference quality $p_0$. For any value of $\mu$, a value of $\sigma$ is found such that the fraction $p_0'$ below $M$ plus the fraction $p_0''$ above $U$ is equal to $p_0$; these values of $\mu$ and $\sigma$ are then taken as the coordinates of a point $\bar{x},s$ on the acceptance boundary. For computational purposes, starting with $\mu$ and finding the proper $\sigma$ is cumbersome, but the parametric equations (111) and (112), which start with an arbitrary division of $p_0$ into two parts $p_0'$ and $p_0''$, are easily used. Given $p_0'$ and $p_0''$, we have

$$\mu - K_{p_0'}\sigma = M \tag{162}$$

and

$$\mu + K_{p_0''}\sigma = U \tag{163}$$

Equations (111) and (112) represent solutions of (162) and (163) for $\mu$ and $\sigma$ in terms of $U$, $M$, $K_{p_0'}$, and $K_{p_0''}$, and thus give the coordinates of a point $\bar{x},s$ on the acceptance boundary.

### 11.2. Approximations in the Basic Formulas

Four incorrect assumptions are made in deriving the preceding formulas:

(1) That $\bar{x} + ks$ is normally distributed; actually, it is not quite normally distributed, since $s$ is not.

(2) That $s$ is distributed about a mean of $\sigma$; actually, it is distributed about a mean of $a\sigma$, where $a < 1$ and is defined by equation (118).

(3) That $s$ is distributed with standard deviation $\sigma/\sqrt{2(N-1)}$; actually, it is distributed with standard deviation of $b\sigma/\sqrt{2(N-1)}$, where $b < 1$ and is defined by equation (119).

(4) That $s$ is distributed with standard deviation $\sigma/\sqrt{2N}$. This assumption applies only to formulas (134) and (135) for $N$.

All these assumptions can be avoided by utilizing the non-central $t$ distribution, tables of which have been published by Johnson and Welch. In fact, the values of $N$ and $k$ and the corresponding true values of $p_1$ and $p_2$ given in Table 1.2 were computed with the aid of tables of the non-central $t$ distribution. The effect of the approximations in the formulas can therefore be examined by comparing the values they give with those given in Table 1.2. Such a comparison exaggerates somewhat the inaccuracy of the formulas because in preparing the table some effort was expended, as is explained in Sec. 11.5, to make the best adjustment to the discontinuous nature of $N$, rather than entering the first values of $N$ and $k$ found. Similar adjustments could be made in using the formulas. Section 4.7 includes a comparison between the results given by the formulas and those in Table 1.2, the comparison being made for the smallest sample size, namely, 6, included in Table 1.2. (It will be noted

that all four of the incorrect assumptions underlying the formulas rapidly become very nearly correct as $N$ increases.)

TABLE 1.8

COMPARISON OF APPROXIMATE AND EXACT VALUES OF $N$ AND $k$

The exact values of $k$ in column (2) have been computed from the non-central $t$ distribution, taking $p_1 = 0.05$, $\alpha = 0.01$, and $N$ as shown in column (1). The exact values of $p_2$ for which $\beta = 0.10$ have then been computed and entered in column (3). From these $p_2$, taking $\beta = 0.10$, $p_1 = 0.05$, and $\alpha = 0.01$, approximate values of $N$ and $k$ have been computed from formulas (10) and (11) or (12). The true values of $p_1$ and $p_2$ for which $\alpha = 0.01$ and $\beta = 0.10$ have been computed from the non-central $t$ distribution.

| Exact | | | Approximate | | | |
|---|---|---|---|---|---|---|
| | $k$ | $p_2$ | $N$ | $k$ | true $p_1$ | true $p_2$ |
| (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| 5 | 0.5445 | 0.5428 | 5 | 0.5150 | 0.0541 | 0.5519 |
| 10 | 0.8037 | 0.3774 | 10 | 0.7856 | 0.0524 | 0.3830 |
| 15 | 0.9292 | 0.3033 | 15 | 0.9163 | 0.0517 | 0.3071 |
| 20 | 1.0083 | 0.2604 | 20 | 0.9983 | 0.0513 | 0.2633 |
| 25 | 1.0643 | 0.2332 | 25 | 1.0560 | 0.0510 | 0.2344 |
| 30 | 1.1069 | 0.2119 | 30 | 1.1000 | 0.0509 | 0.2137 |
| 35 | 1.1409 | 0.1966 | 35 | 1.1348 | 0.0507 | 0.1981 |
| 40 | 1.1688 | 0.1846 | 40 | 1.1633 | 0.0507 | 0.1859 |
| 45 | 1.1922 | 0.1748 | 45 | 1.1874 | 0.0506 | 0.1759 |
| 50 | 1.2125 | 0.1666 | 50 | 1.2082 | 0.0505 | 0.1676 |
| 65 | 1.2592 | 0.1489 | 64 | 1.2556 | 0.0501 | 0.1500 |
| 75 | 1.2828 | 0.1404 | 74 | 1.2797 | 0.0501 | 0.1413 |
| 100 | 1.3264 | 0.1256 | 99 | 1.3241 | 0.0501 | 0.1262 |

Table 1.8 shows a few comparisons of the values of $k$ and $N$ given by formulas (132) and (134) or (135) with the exact values. In these cases $\alpha$ has been fixed at 0.01 and $p_1$ at 0.05. Various values of $N$ have been selected and the resulting lot quality $p_2$ for which $\beta = 0.10$ has been calculated from the non-central $t$ distribution. Thus the exact values of $N$ and $k$ result in an OC curve passing exactly through the points defined by $p_1, \alpha$ and $p_2, \beta$. If $p_1, \alpha$ and $p_2, \beta$ are both selected arbitrarily, it is not usually possible to find values of $N$ and $k$ such that the OC curve passes exactly through both points. Thus the exact values of $N$ and $k$ given in Table 1.8 are more exact than could ordinarily be achieved in practical problems, while those given by the formulas are characteristic of the accuracy realizable in practice. Columns (6) and (7) of the table show the true values of $p_1$ and $p_2$ for which $\alpha = 0.01$ and $\beta = 0.10$ when the approximate $N$ and $k$ are used. The differences between 0.05 and the values in column (6) and between the values in columns (3) and (7)

are a measure of the effect of the approximations. An alternative measure would be the error in the true $\alpha$ and $\beta$ at the intended $p_1$ and $p_2$; this error would probably be smaller than that in $p_1$ and $p_2$ because the OC curves are usually fairly flat in the neighborhood of $p_1$ and $p_2$. While it cannot be claimed that Table 1.8 gives a representative coverage of all kinds of plans, it can be claimed that, for those comparisons given, the discrepancy between the approximate and exact results is maximized.

### 11.3. Computational Formulas

The use of the acceptance criterion

$$\bar{x} + ks \leq U \tag{164}$$

requires the calculation of the sample mean $\bar{x}$ and the standard deviation $s$. The acceptance criterion can, however, be expressed in terms of simpler functions of the observations, namely,

$$X = \Sigma x \quad \text{(sum of the observations)} \tag{165}$$
$$Y = \Sigma x^2 \quad \text{(sum of the squares of the observations)} \tag{166}$$

If $k$ is positive, as it almost invariably will be, the lot may be rejected without regard to $Y$ if

$$X > NU \tag{167}$$

In this case, $\bar{x}$ exceeds $U$, so since $s$ cannot be negative, $\bar{x} + ks$ exceeds $U$.

The values of $\bar{x}$ and $s$ for a sample of $N$ observations can be calculated from

$$\bar{x} = \frac{\Sigma x}{N} \tag{168}$$

and

$$s = \sqrt{\frac{N\Sigma x^2 - (\Sigma x)^2}{N(N-1)}} \tag{169}$$

In terms of $X$ and $Y$, equations (168) and (169) can be written

$$\bar{x} = \frac{X}{N} \tag{170}$$

and

$$s = \sqrt{\frac{NY - X^2}{N(N-1)}} \tag{171}$$

Replacing $\bar{x}$ and $s$ in (164) by their values from (170) and (171) gives

$$\frac{X}{N} + k\sqrt{\frac{NY - X^2}{N(N-1)}} \leq U \tag{172}$$

Subtracting $X/N$ from both sides of inequality (172) and squaring gives

$$\frac{NY - X^2}{N(N-1)} \cdot k^2 \leq U^2 - \frac{2XU}{N} + \frac{X^2}{N^2} \tag{173}$$

Clearing fractions and rearranging terms gives

$$Y \leq aX^2 - bX + c \qquad (174)$$

where

$$a = \frac{Nk^2 + N - 1}{N^2 k^2} \qquad (175)$$

$$b = \frac{2U(N - 1)}{Nk^2} \qquad (176)$$

$$c = \frac{U^2(N - 1)}{k^2} \qquad (177)$$

If inequality (174) is fulfilled, the lot is accepted; if inequality (174) is not fulfilled, the lot is rejected.

Using the equality sign in (174) gives the equation of a parabola

$$Y_R = aX^2 - bX + c \qquad (178)$$

which shows, for each value of $X$, the maximum value of $Y$ permitting acceptance, that is, the upper boundary of the acceptance region or the lower boundary of the rejection region.

The minimum value of the curve $Y_R$ gives a value of $Y$ below which the lot is acceptable without regard to $X$. Differentiating (178) with respect to $X$ gives

$$\frac{dY_R}{dX} = 2aX - b \qquad (179)$$

Setting $\dfrac{dY_R}{dX}$ equal to zero and solving for $X$ gives

$$\check{X} = \frac{b}{2a} \qquad (180)$$

Since

$$\frac{d^2 Y_R}{dX^2} = 2a > 0 \qquad (181)$$

$\check{X}$ is the value of $X$ at which the minimum value $\check{Y}_R$ of $Y_R$ occurs. Replacing $X$ in (178) by the value of $\check{X}$ given by (180) gives, after simplification,

$$\check{Y}_R = c - \frac{b^2}{4a} \qquad (182)$$

Substituting the values of $a$, $b$, and $c$ given by (175), (176), and (177) into (180) and (182), and simplifying, gives

$$\check{X} = \frac{UN(N - 1)}{Nk^2 + N - 1} \qquad (183)$$

$$\check{Y}_R = \frac{U^2 N(N - 1)}{Nk^2 + N - 1} \qquad (184)$$

The coordinates of the minimum point, $X = \check{X}$ and $Y = \check{Y}_R$, can be

approximated by replacing by $N$ the $N - 1$ in the denominators of (183) and (184)

$$\text{Approximate } \check{X} = \frac{U(N-1)}{k^2 + 1} \tag{185}$$

$$\text{Approximate } \check{Y}_R = \frac{U^2(N-1)}{k^2 + 1} \tag{186}$$

These approximate values are smaller than the exact values since they were obtained by increasing the denominators; if, therefore,

$$Y \leq \frac{U^2(N-1)}{k^2 + 1} \tag{187}$$

$Y$ will certainly be less than the minimum value of $Y_R$, and the lot may be accepted without regard to $X$.

The acceptance region has a lower boundary corresponding, for each value of $X$, with the minimum value of $Y$ algebraically possible. The sum of the squares of the deviations of the observations from their mean can be written

$$\Sigma(x - \bar{x})^2 = \Sigma x^2 - \frac{(\Sigma x)^2}{N} \tag{188}$$

or, in terms of $X$ and $Y$

$$\Sigma(x - \bar{x})^2 = Y - \frac{X^2}{N} \tag{189}$$

Since $\Sigma(x - \bar{x})^2$ cannot be negative,

$$Y \geq \frac{X^2}{N} \tag{190}$$

For any given sum of observations $X$, the value of $Y$ will be $\dfrac{X^2}{N}$ when the sum of the squares of the deviations of the observations from their mean is zero; that is, when all the observations are equal. The lower boundary of the acceptance region, or upper boundary of the error region, is, therefore,

$$Y_E = \frac{X^2}{N} \tag{191}$$

A sample point lying below this line would indicate a computational error.

By equating $Y_R$ and $Y_E$ given by (178) and (191), it can be shown that the lower and upper boundaries of the acceptance region meet at the point $X = NU$ and $Y = NU^2$. The two curves are tangent at this point. Beyond it, they again enclose an area similar to the acceptance region; this area beyond the point of tangency would be the acceptance region for a test in which $k$ is negative.

The graphical procedure of Sec. 7.3.1 consists of drawing on a chart the upper and lower boundaries of the acceptance region, given by (178) and (191); a point representing the sample is plotted on this chart and the lot accepted or rejected according as the point lies within or above the acceptance region. This graphical procedure is impractical, however, when the observations are large. In this case, the range of variation in $Y$ as $X$ varies is large relative to the width of the acceptance region; and the upper and lower boundaries of the acceptance region are too close for most practical purposes.

An adjustment of the observations can be made (a) to overcome this difficulty in the graphical procedure or (b) to find a criterion more effective than formula (187) in permitting immediate acceptance. The adjustment consists of subtracting an adjustment constant $A$ from each of the observations, $A$ being chosen according to the purpose (a) or (b) for which the adjustment is being used. It is not necessary to adjust each observation; it is necessary only to compute the sum of the squares of the adjusted observations $Y'$ from $X$ and $Y$ by

$$Y' = \Sigma(X - A)^2 \tag{192}$$
$$= Y - 2AX + NA^2$$

Adjusting $Y_R$ and $Y_E$, given in (178) and (191), by (192) (that is, subtracting $2AX$ and adding $NA^2$) gives the upper and lower boundaries of the acceptance region

$$Y'_R = aX^2 - (b + 2A)X + (c + NA^2) \tag{193}$$
$$Y'_E = \frac{X^2}{N} - 2AX + NA^2 \tag{194}$$

where $a$, $b$, and $c$ are defined by (175), (176), and (177).

It can be shown that the minimum value $\check{Y}'_R$ of $Y'_R$ is

$$\check{Y}'_R = \frac{(U - A)^2 N(N - 1)}{Nk^2 + N - 1} \tag{195}$$

Replacing by $N$ the $N - 1$ in the denominator of $\check{Y}'_R$ gives the simplified criterion for immediate acceptance

$$Y' \leq \frac{(U - A)^2(N - 1)}{k^2 + 1} \tag{196}$$

If the graphical procedure is used, the adjustment constant $A$ should be chosen so as to minimize the variation in $Y$ as $X$ varies over the region of likely sample values. This flattens the curves defining the acceptance region, thus enabling the acceptance region to encompass a larger portion of the chart's area and thereby improving the accuracy of

plotting. To determine the optimum value of $A$, consider the slope of $Y'_E$ at a given value of $X$

$$\frac{dY'_E}{dX} = \frac{2X}{N} - 2A \tag{197}$$

For a given value of $X$, the slope is zero when $A = X/N$. The optimum value of $A$ depends, therefore, on the observed value of $X$. If before the sample is inspected $\mu'$ is the best estimate of the mean of the characteristic being inspected, the optimum value of $A$, for graphical purposes, is

$$A = \mu' \tag{198}$$

since the expected value of $X$ is $N\mu'$.

If the graphical procedure is not used, the adjustment constant $A$ should be chosen so as to maximize the effectiveness of formula (196) in permitting immediate acceptance. This involves finding the value of $A$ that minimizes the $Y$ corresponding to the minimum value $Y' = \check{Y}'_R$. Differentiating (193) with respect to $X$ gives

$$\frac{dY'_R}{dX} = 2aX - (b + 2A) \tag{199}$$

Since

$$\frac{d^2Y'_R}{dX^2} = 2a > 0 \tag{200}$$

setting $\dfrac{dY'_R}{dX}$ equal to zero and solving for $X$ gives the value of $X$ at which $Y'_R$ is a minimum, namely,

$$X = \frac{b + 2A}{2a} \tag{201}$$

The value of $Y'_R$ calculated from (193) for the point whose abscissa is given by (201) is the minimum value of $Y'_R$

$$\check{Y}'_R = -\frac{(b + 2A)^2}{4a} + (c + NA^2) \tag{202}$$

The value of $\check{Y}'_R$ given by (195) can be obtained from (202) by replacing $a$, $b$, and $c$ by their values given by (175), (176), and (177). Finding the value of $Y$ corresponding to $Y' = \check{Y}'_R$, that is, adding $2AX$ to, and subtracting $NA^2$ from, (202) gives

$$Y = -\frac{(b + 2A)^2}{4a} + (c + NA^2) + 2AX - NA^2 \tag{203}$$

$$= -\frac{(b + 2A)^2}{4a} + c + 2AX$$

Now, the optimum value of $A$ is that value which makes the $Y$ corresponding to the minimum value $\check{Y}'_R$ as large as possible, since, for a given

value of $X$, if $Y$ is less than the value given by (203), the lot is accepted. Differentiating (203) with respect to $A$ gives

$$\frac{dY}{dA} = -\frac{(b + 2A)}{a} + 2X \tag{204}$$

Since

$$\frac{d^2Y}{dA^2} = -\frac{2}{a} < 0 \tag{205}$$

setting $\frac{dY}{dA}$ equal to zero and solving for $A$ gives the value of $A$ at which $Y$ is a maximum, namely,

$$A = aX - \frac{b}{2} \tag{206}$$

The value of $A$ given in (206) is a function of $X$, so the optimum value of $A$ depends on the observed value of $X$. If before the sample is inspected $\mu'$ is the best estimate of $\bar{x}$, the anticipated value of $X$ is $N\mu'$. Replacing $X$ in (206) by its anticipated value gives

$$A = aN\mu' - \frac{b}{2} \tag{207}$$

Replacing $a$ and $b$ in (207) by their values given by (175) and (176), and simplifying, gives

$$A = \frac{(Nk^2 + N - 1)\mu' - (N - 1)U}{Nk^2} \tag{208}$$

Replacing by $N$ the $N - 1$ in the numerator of (208) gives

$$\text{Approximate } A = \frac{(k^2 + 1)\mu' - U}{k^2} \tag{209}$$

Using the value of $A$ given by either (208) or (209) makes (196) very nearly the most effective in permitting immediate acceptance. In terms of $Y$, replacing the $A$ in (203) by (207), and simplifying, gives the linear condition for acceptance

$$Y \leq c' - b'X \tag{210}$$

where

$$b' = b - 2N\mu'a \tag{211}$$
$$c' = c - N^2\mu'^2a \tag{212}$$

and $a$, $b$, and $c$ are defined by (175), (176), and (177).

## 11.4. Use of Non-central $t$ Distribution

### 11.4.1. Introduction.—The acceptance criterion

$$\bar{x} + ks \leq U \tag{213}$$

can be written

$$\frac{U - \bar{x}}{s} \geq k \tag{214}$$

The quantity $(U - \bar{x})/s$ is a function of the sample data and $k$ is a constant; using the acceptance criterion (214) is thus equivalent to comparing the value of the sample statistic $(U - \bar{x})/s$ with a minimum level for acceptance $k$.

If many samples of $N$ observations are chosen at random from the same population (or lot), the values of $(U - \bar{x})/s$ obtained from the various samples will vary. The proper determination of $k$ requires a knowledge of the sampling distribution of $(U - \bar{x})/s$.

The distribution of the quantity $t$, defined as

$$t = \sqrt{N} \left( \frac{U - \bar{x}}{s} \right) \tag{215}$$

is known as the non-central $t$ distribution, of which the familiar Student's $t$ distribution is a special case corresponding with $Ex = U$. Tables of the non-central $t$ distribution with an explanation of their application to the problems considered in this chapter have been given by Johnson and Welch.

**11.4.2. Notation Used in the Tables.**—It will be necessary to use the notation of Johnson and Welch to explain the application of their tables to the problems discussed in this chapter. The following list indicates the changes in notation:*

|  | This chapter | Johnson and Welch |
|---|---|---|
| Value of variable that defines quality............ | $U$ | $L$ |
| Sample statistic.............................. | $k_o = \dfrac{U - \bar{x}}{s}$ | $u = \dfrac{L - \bar{x}}{s}$ |
| Acceptable level of statistic................... | $k$ | $u_0$ |
| Normal deviate corresponding with proportion $p$ defective in lot........................... | $K_p$ | $U$ |
| Normal deviate corresponding with probability $\epsilon$ of accepting a lot........................... | $K_\epsilon$ | $K$ |

* A word of explanation is in order about our departure from the Johnson-Welch and Jennett-Welch notation, since their notation is satisfactory and established. Throughout all three of our volumes the letter $L$ is consistently used to represent the ordinate of an OC curve; as a substitute $U$ has the mnemonic advantage of association with "upper limit." It seems unnecessary to use two symbols for the normal deviate, since the type of subscript (Greek letter, numeral, and so forth) can discriminate between normal deviates defining quality and those describing operating characteristics. For the sample statistic, $k$ accords with the lot parameter $K$. Logic suggests a subscript for the acceptable level of the statistic, and Johnson and Welch use $o$; but because the acceptable level enters the analysis so frequently and other values so seldom, we use no subscript. Logic also suggests a subscript for a particular observed value of the statistic, and we use $o$.

In the notation of Johnson and Welch, the acceptance criterion (214) becomes

$$u \geq u_0 \tag{216}$$

Let $z$ be a quantity distributed normally about zero with unit standard deviation, and let $w$ be a quantity distributed independently of $z$ and in the same way as an estimate of variance computed from a sample from a normal population containing $f$ degrees of freedom. That is, $w$ is distributed as $\chi^2/f$ with $f$ degrees of freedom. If $t$ is defined by

$$t = \frac{z + \delta}{\sqrt{w}} \tag{217}$$

where $\delta$ is some constant, then $t$ has the non-central $t$ distribution, depending only on the eccentricity $\delta$ and the number of degrees of freedom $f$.

To see that the statistic $t$ is closely related to the non-central $t$ distribution, consider the distribution of values of $u$ from random samples of $N$ observations from a normal population with mean $\mu$ and standard deviation $\sigma$. Suppose that exactly a proportion $p$ of the items in the population exceed a given limit $L$ and that $U$ is the standardized normal deviate exceeded with probability $p$, that is,

$$\frac{1}{\sqrt{2\pi}} \int_U^{\infty} e^{-\frac{r^2}{2}} \, dr = p \tag{218}$$

Let $\bar{x}$ and $s$ be the sample mean and standard deviation, respectively. Then, in the notation of Johnson and Welch, $\sqrt{N}\, u$ can be written as (217) with

$$z = \frac{- \sqrt{N}(\bar{x} - u)}{\sigma} \tag{219}$$

$$w = \frac{s^2}{\sigma^2} \tag{220}$$

$$f = N - 1 \tag{221}$$

$$\delta = \sqrt{N}\, U \tag{222}$$

Hence, $\sqrt{N}\, u$ is distributed as a non-central $t$ with $f = N - 1$ and $\delta = \sqrt{N}\, U$.

The distribution of non-central $t$ requires a triple-entry table, since the probability that $t$ exceeds a given level $t_0$ depends on $f$, $\delta$, and $t_0$. This probability is denoted by $P(f, \delta, t_0)$. Johnson and Welch also use $t(f, \delta, \epsilon)$ to denote that value of $t_0$ for which the probability $P(f, \delta, t_0)$ takes a specified value $\epsilon$. The value of $\delta$ that makes $P(f, \delta, t_0) = \epsilon$ is denoted by $\delta(f, t_0, \epsilon)$.

**11.4.3. Description of the Tables.**—The problems for which the non-central $t$ tables are used in this chapter require the evaluation of either

$\delta(f,t_0,\epsilon)$ or $t(f,\delta,\epsilon)$. Table IV given by Johnson and Welch facilitates the direct calculation of these for 17 probability levels corresponding with $\epsilon$ and $1-\epsilon$ equal to 0.005, 0.01, and 0.025, 0.05, 0.1, 0.2, 0.3, 0.4, and 0.5.

In Table IV, a quantity $\lambda = \lambda(f,t_0,\epsilon)$ is given such that

$$\delta(f,t_0,\epsilon) = t_0 - \lambda\sqrt{1 + \frac{t_0^2}{2f}} \tag{223}$$

and

$$t(f,\delta,\epsilon) = \frac{\delta + \lambda\sqrt{1 + \dfrac{\delta^2}{2f} - \dfrac{\lambda^2}{2f}}}{1 - \dfrac{\lambda^2}{2f}} \tag{224}$$

Equations (42) and (21) of this chapter are essentially approximations for equations (223) and (224), respectively, with $\lambda$ in (223) and (224) replaced by $K_\epsilon$, the normal deviate exceeded with probability $\epsilon$. Even for small samples, it is found that $\lambda$ does not differ much from $K_\epsilon$.

Table IV consists of nine double-entry panels (one for each of the values of $\epsilon$ listed above) showing $\lambda$ as a function of $f$ and $t_0$. Values of $\lambda$ are given corresponding to $f = 4, 5, 6, 7, 8, 9, 16, 36, 144$, and $\infty$. In order to simplify interpolation for intermediate values of $f$ greater than 9, the values of $f = 9, 16, 36, 144$ and $\infty$ are such that $12/\sqrt{f} = 4, 3, 2, 1$, and 0, respectively; that is, $\lambda$ is tabled at equal intervals of $12/\sqrt{f}$. The whole range of $t_0$ from $-\infty$ to $\infty$ is covered as follows: For $t_0/\sqrt{2f}$ between $-\infty$ and $-0.75$ or between $+0.75$ and $+\infty$, $\lambda$ is tabled against the quantity

$$y = \frac{1}{\sqrt{1 + \dfrac{t_0^2}{2f}}} \tag{225}$$

For $t_0/\sqrt{2f}$ between $-0.75$ and $+0.75$, $\lambda$ is tabled against

$$y' = \frac{\dfrac{t_0}{\sqrt{2f}}}{\sqrt{1 + \dfrac{t_0^2}{2f}}} \tag{226}$$

**11.4.4. Calculation of $\delta(f,t_0,\epsilon)$.**—For given $f$, $t_0$, and $\epsilon \leq 0.5$, the steps required for the calculation of $\delta(f,t_0,\epsilon)$ are given below. For $\epsilon > 0.5$, the relation

$$\delta(f,t_0,\epsilon) = -\delta(f,-t_0,1-\epsilon) \tag{227}$$

is used. That is, if $\delta(9,12,0.95)$ is required, calculate $\delta(9,-12,0.05)$ by the procedure outlined below and take the negative of this value as the required value of $\delta(9,12,0.95)$.

*Step* 1.   Calculate $t_0/\sqrt{2f}$.

*Step* 2.   Find $y = \dfrac{1}{\sqrt{1 + t_0^2/2f}}$ or $y' = \dfrac{t_0/\sqrt{2f}}{\sqrt{1 + t_0^2/2f}}$ according as $\left|\dfrac{t_0}{\sqrt{2f}}\right|$ is greater than or less than 0.75.

*Step* 3.   If $f$ is greater than 9, calculate $12/\sqrt{f}$.

*Step* 4.   From the panel of Table IV corresponding with the given $\epsilon$, obtain $\lambda = \lambda(f,t_0,\epsilon)$ by interpolating with respect to the quantities obtained in steps 2 and 3.

*Step* 5.   Calculate $\delta(f,t_0,\epsilon) = t_0 - \lambda\sqrt{1 + t_0^2/2f}$.

**11.4.5. Calculation of $t(f,\delta,\epsilon)$.**—The calculation of $t(f,\delta,\epsilon)$ is accomplished by an iterative method, since $\lambda$ is given in Table IV as a function of $t_0$. The successive steps in the iterative method for calculating $t(f,\delta,\epsilon)$ for given $f$, $\delta$, and $\epsilon \leq 0.5$ are given below. For $\epsilon > 0.5$, the relation

$$t(f,\delta,\epsilon) = -t(f,-\delta,1-\epsilon) \tag{228}$$

is used. That is, if $t(9,5,0.90)$ is required, calculate $t(9,-5,0.10)$ by the procedure outlined below, and take the negative of this value as the required $t(9,5,0.90)$.

*Step* 1.   Determine $K_\epsilon$, the normal deviate exceeded with probability $\epsilon$, which can be found in Table 1.1.

*Step* 2.   Calculate the first approximation $t_1$ from

$$t_1 = \frac{\delta + K_\epsilon\sqrt{1 + \dfrac{\delta^2}{2f} - \dfrac{K_\epsilon^2}{2f}}}{1 - \dfrac{K_\epsilon^2}{2f}}$$

*Step* 3.   Calculate $t_1/\sqrt{2f}$.

*Step* 4.   Find $y = \dfrac{1}{\sqrt{1 + t_1^2/2f}}$ or $y' = \dfrac{t_1/\sqrt{2f}}{\sqrt{1 + t_1^2/2f}}$ according as $\left|\dfrac{t_1}{\sqrt{2f}}\right|$ is greater than or less than 0.75.

*Step* 5.   If $f$ is greater than 9, calculate $12/\sqrt{f}$.

*Step* 6.   From the panel of Table IV corresponding with the given $\epsilon$, obtain $\lambda_1 = \lambda(f,t_1,\epsilon)$ by interpolating with respect to the quantities obtained in steps 4 and 5.

*Step* 7.   Calculate a second approximation $t_2$ from

$$t_2 = \frac{\delta + \lambda_1\sqrt{1 + \dfrac{\delta^2}{2f} - \dfrac{\lambda_1^2}{2f}}}{\left(1 - \dfrac{\lambda_1^2}{2f}\right)}$$

*Step* 8.   Repeat steps 3 through 7, replacing $t_1$ by $t_2$, and calculate a third approximation $t_3$.  Repeat these steps until two successive approximations are the same; usually not more than two approximations are required.  Take for $t(f,\delta,\epsilon)$ the value so obtained.

### 11.4.6. Illustration of the Use of the Non-central $t$ Tables in Specific Problems

*Problem* I.   *Given:* $N$, $p_1$, $\epsilon$  (in our notation, $N$, $p_1$, $1 - \alpha$ or
$\qquad\qquad\qquad\qquad N$, $p_2$, $\beta$)
$\qquad$*Required:* $u_0$  (in our notation, $k$)

A single-sampling plan based on variables is defined by a sample size $N$ and a value $u_0$ of the statistic $u$.  If $N$ is fixed, $u_0$ can be determined so that the operating-characteristic curve for the plan passes through a specified point $(p_1,\epsilon)$; that is, for a fixed sample size, a plan is chosen for which the probability of accepting lots of quality $p_1$ is $\epsilon$.

For given $N$, $p_1$, and $\epsilon$, the value $u_0$ that defines the plan is given, in the notation of Johnson and Welch, by

$$\sqrt{N}u_0 = t(N - 1,\sqrt{N}U_1,\epsilon) \qquad (229)$$

where $U_1$ is the normal deviate exceeded with probability $p_1$.  Thus, for a given sample size $N$, quality (fraction defective) $p_1$, and probability of acceptance $\epsilon$, the determination of $u_0$ requires the calculation of $t(f,\delta,\epsilon)$ where $f = N - 1$ and $\delta = \sqrt{N}U_1$.

An outline of the method for determining $u_0$ for given values of $N$, $p_1$ and $\epsilon$ is given below.

*Given:*   $N$ = sample size
$\qquad\qquad p_1$ = a proportion defining quality
$\qquad\qquad \epsilon$ = desired probability of accepting lots of quality $p_1$

*Step* 1.   Determine $U_1$, the normal deviate exceeded with probability $p_1$, from Table 1.1.
*Step* 2.   Calculate
$$f = N - 1$$
$$\delta = \sqrt{N}U_1$$

*Step* 3.   Calculate $t(f,\delta,\epsilon)$ by the method described in Sec. 11.4.5.
*Step* 4.   Calculate $u_0 = t(f,\delta,\epsilon)/\sqrt{N}$.

*Problem* II.   *Given:* $N$, $u_0$, $\epsilon$  (in our notation, $N$, $k$, $L_p$)
$\qquad\qquad\qquad$*Required:* $p$

For every single-sampling plan (defined by $N$ and $u_0$), there is an operating-characteristic (OC) curve which shows the probability $\epsilon$ of

accepting a lot of quality $p$.    From the Johnson and Welch tables we can compute the values of $p$ corresponding to $\epsilon$ or $1 - \epsilon$ equal to 0.005, 0.01, 0.025, 0.05, 0.1, 0.2, 0.3, 0.4, or 0.5.    This requires finding $\delta(f,t_0,\epsilon)$ where $f = N - 1$ and $t_0 = \sqrt{N}u_0$.    The value of $U$ corresponding to this value of $\delta$ is $\delta/\sqrt{N}$.    The value of $p$ corresponding to the given $\epsilon$ is then the probability that a normal deviate will exceed this $U$, which may be found from Table 1.5.

*Step* 1.    Calculate

$$t_0 = \sqrt{N}u_0$$
$$f = N - 1$$

*Step* 2.    Calculate $\delta(f,t_0,\epsilon)$ by the method described in Sec. 11.4.4.

*Step* 3.    Calculate $U = \delta(f,t_0,\epsilon) / \sqrt{N}$.

*Step* 4.    Obtain the required $p$ by entering Table 1.5 with $K$ equal to the value of $U$ found in step 3.

**11.4.7.  Work Sheets.**— Suppose that the sample size is fixed at 30, and a plan is required for which a lot having 0.1 percent defective will be accepted with probability 0.99.    That is, $N = 30$, $p_1 = 0.001$, and $\epsilon = 0.99$.    Work sheets 1, 2, and 3 (Tables 1.9, 1.10, and 1.11) represent the calculations necessary to determine $u_0$; work sheet 4 (Table 1.12) represents those for the OC curve.    The numerical entries have been rounded from calculations originally carried to more decimal places, so they do not always agree exactly with operations carried out on the 5-decimal-place figures shown.

TABLE 1.9

WORK SHEET 1 FOR DETERMINING $u_0$ (IN OUR NOTATION $k$) FROM NON-CENTRAL $t$ TABLES

| Row number (1) | Operation (2) | Result (3) | Example (4) |
|---|---|---|---|
| (1) | Given | $p_1$ | 0.001 |
| (2) | Given | $\epsilon$ | 0.99 |
| (3) | Given | $N$ | 30. |
| (4) | Table 1.1 | $U_1$ | 3.09023 |
| (5) | $\sqrt{(3)}$ | $\sqrt{N}$ | 5.47723 |
| (6) | $(3) - 1$ | $f = N - 1$ | 29. |
| (7) | $(4) \times (5)$ | $\delta = \sqrt{N}\,U_1$ | 16.92590 |
| (8) | Sheet 2, row (14), last column | $t(f,\delta,\epsilon)$ | 12.47223 |
| (9) | $(8) \div (5)$ | $u_0 = \dfrac{t(f,\delta,\epsilon)}{\sqrt{N}}$ | 2.27711 |

TABLE 1.10

WORK SHEET 2 FOR DETERMINING $u_0$—COMPUTATION OF $t(f, \delta, \epsilon)$

| Row number (1) | Operation (2) | Result (3) | First approximation (4) | Second approximation (5) |
|---|---|---|---|---|
| (1) | Sheet 1, line 6 | $f$ | 29 | ← |
| (2) | Sheet 1, line 7* | $\delta$ or $-\delta$ | −16.92590 | ← |
| (3) | Sheet 1, line 2* | $\epsilon$ or $1 - \epsilon$ | 0.01 | ← |
| (4) | 2 × (1) | $2f$ | 58 | ← |
| (5) | (2)² | $\delta^2$ | 286.48606 | ← |
| (6) | (5) ÷ (4) | $\dfrac{\delta^2}{2f}$ | 4.93941 | ← |
| (7) | Table 1.1 | $K_\epsilon$ | 2.32635 | 2.321† |
| (8) | (7)² | $K_\epsilon^2$ | 5.41190 | 5.38704 |
| (9) | (8) ÷ (4) | $\dfrac{K_\epsilon^2}{2f}$ | 0.09331 | 0.09288 |
| (10) | 1 + (6) − (9) | $1 + \dfrac{\delta^2}{2f} - \dfrac{K_\epsilon^2}{2f}$ | 5.84611 | 5.84654 |
| (11) | $\sqrt{(10)}$ | $\sqrt{1 + \dfrac{\delta^2}{2f} - \dfrac{K_\epsilon^2}{2f}}$ | 2.41787 | 2.41796 |
| (12) | (2) + (7) × (11) | $\delta + K_\epsilon \sqrt{1 + \dfrac{\delta^2}{2f} - \dfrac{K_\epsilon^2}{2f}}$ | −11.30109 | −11.31381 |
| (13) | 1 − (9) | $1 - \dfrac{K_\epsilon^2}{2f}$ | 0.90669 | 0.90712 |
| (14) | (12) ÷ (13) | $t = t(f, \delta, \epsilon)$ | −12.46410 | −12.47223 |
| (15) | $\sqrt{(4)}$ | $\sqrt{2f}$ | 7.61577 | ← |
| (16) | (14) ÷ (15) | $\dfrac{t}{\sqrt{2f}}$ | −1.63662 | −1.63768 |
| (17) | (14)² | $t^2$ | 155.35379 | 155.55660 |
| (18) | (17) ÷ (4) | $\dfrac{t^2}{2f}$ | 2.67851 | 2.68201 |
| (19) | 1 + (18) | $1 + \dfrac{t^2}{2f}$ | 3.67851 | 3.68201 |
| (20) | $\sqrt{(19)}$ | $\sqrt{1 + \dfrac{t^2}{2f}}$ | 1.91795 | 1.91886 |
| (21) | 1 ÷ (20) | $y = 1 \Big/ \sqrt{1 + \dfrac{t^2}{2f}}$ | 0.52139 | 0.52114 |
| (22) | (16) ÷ (20) | $y' = \dfrac{t}{\sqrt{2f}} \Big/ \sqrt{1 + \dfrac{t^2}{2f}}$ | ‡ | ‡ |
| (23) | $\sqrt{(1)}$ | $\sqrt{f}$ | 5.38516 | ← |
| (24) | 12 ÷ (23) | $\dfrac{12}{\sqrt{f}}$ | 2.22834 | ← |
| (25) | Sheet 3, row (17), last column | $\lambda$ | 2.321 | 2.321 |

* Since $\epsilon > 0.5$, we change the sign of $\delta$ and replace $\epsilon$ by $1 - \epsilon$.

† The figure in row (7), column (5) is the value of $\lambda_1$ given in row (25), column (4) and is used in place of $K_\epsilon$ in the succeeding calculations for $\lambda_2$.

‡ No figures are given for $y'$ in this example because $t_0/\sqrt{2f} > 0.75$.

TABLE 1.11
WORK SHEET 3 FOR DETERMINING $u_0$—LINEAR INTERPOLATION FOR $\lambda(a,b)$

| Row number | Operation | Result | First approximation | Second approximation |
|---|---|---|---|---|
| (1) | (2) | (3) | (4) | (5) |
| (1) | Sheet 2, line 21* | $a$ | 0.52139 | 0.52114 |
| (2) | Sheet 2, line 24* | $b$ | 2.2283 | 2.2283 |
| (3) | Table IV* | $a_1$ | 0.5 | 0.5 |
| (4) | Table IV* | $a_2$ | 0.6 | 0.6 |
| (5) | Table IV* | $b_1$ | 3. | 3. |
| (6) | Table IV* | $b_2$ | 2. | 2. |
| (7) | $\dfrac{(3)-(1)}{(3)-(4)}$ | $c_1 = \dfrac{a_1-a}{a_1-a_2}$ | 0.2139 | 0.2114 |
| (8) | $\dfrac{(1)-(4)}{(3)-(4)}$ | $c_2 = \dfrac{a-a_2}{a_1-a_2}$ | 0.7861 | 0.7886 |
| (9) | $\dfrac{(5)-(2)}{(5)-(6)}$ | $d_1 = \dfrac{b_1-b}{b_1-b_2}$ | 0.7717 | 0.7717 |
| (10) | $\dfrac{(2)-(6)}{(5)-(6)}$ | $d_2 = \dfrac{b-b_2}{b_1-b_2}$ | 0.2283 | 0.2283 |
| (11) | Table IV* | $\lambda_{11}$ | 2.318 | ← |
| (12) | Table IV* | $\lambda_{12}$ | 2.324 | ← |
| (13) | Table IV* | $\lambda_{21}$ | 2.310 | ← |
| (14) | Table IV* | $\lambda_{22}$ | 2.318 | ← |
| (15) | (8)(11) + (7)(13) | $c_2\lambda_{11} + c_1\lambda_{21}$ | 2.31629 | 2.31631 |
| (16) | (8)(12) + (7)(14) | $c_2\lambda_{12} + c_1\lambda_{22}$ | 2.32272 | 2.32273 |
| (17) | (10)(15) + (9)(16) | $\lambda(ab)$ $= d_2(c_2\lambda_{11} + c_1\lambda_{21})$ $+ d_1(c_2\lambda_{12} + c_1\lambda_{22})$ | 2.321 | 2.321 |

* For the given $\epsilon$, $\lambda$ is a function of $f$ and $t_0$ and is given in Johnson and Welch's Table IV, for specified values of $y$ (or $y'$) and $f$ (or $12/\sqrt{f}$). $a$ and $b$ are the values of the arguments (in this case $y$ and $12/\sqrt{f}$, respectively) for which $\lambda$ is required. $a_1$, $a_2$, $b_1$, and $b_2$ are the values of the arguments for which Table IV gives $\lambda$ values and between which we interpolate to find $\lambda(a,b)$. $\lambda_{11}$, $\lambda_{12}$, $\lambda_{21}$, and $\lambda_{22}$ are the tabular values corresponding to $a_1$, $a_2$, $b_1$, and $b_2$. The following scheme explains the notation;

| | ... | $b_1$ | $b_2$ | ... |
|---|---|---|---|---|
| ... $a_1$ $a_2$ ... | | $\lambda_{11}$ $\lambda_{21}$ | $\lambda_{12}$ $\lambda_{22}$ | |

The computation of the OC curve for the above plan is illustrated by finding the values of $p$ for which $L_p$ is 0.2 and 0.7. The figures given in Table 1.12 represent the calculations necessary to determine the values of $p$ corresponding with these two values of $L_p$. Additional columns would be required for additional points on the OC curve.

TABLE 1.12

WORK SHEET 4 FOR OPERATING-CHARACTERISTIC CURVE

| (1) Row number | (2) Operation | (3) Result | (4) First point | (5) Second point |
|---|---|---|---|---|
| (1) | Given | $L_p$ | 0.2 | 0.7 |
| (2) | Smaller of $L_p$ and $1 - L_p$ | $\epsilon$ | 0.2 | 0.3 |
| (3) | Sheet 3, line 1, last column | $a$ | 0.52114 | ← |
| (4) | Sheet 3, line 2, last column | $b$ | 2.2283 | ← |
| (5) | Sheet 3, row 3* | $a_1$ | 0.6 | 0.5 |
| (6) | Sheet 3, row 4* | $a_2$ | 0.5 | 0.6 |
| (7) | Sheet 3, row 5 | $b_1$ | 3. | ← |
| (8) | Sheet 3, row 6 | $b_2$ | 2. | ← |
| (9) | $\dfrac{(5) - (3)}{(5) - (6)}$ | $c_1 = \dfrac{a_1 - a}{a_1 - a_2}$ | 0.7886 | 0.2114 |
| (10) | $\dfrac{(3) - (6)}{(5) - (6)}$ | $c_2 = \dfrac{a - a_2}{a_1 - a_2}$ | 0.2114 | 0.7886 |
| (11) | $\dfrac{(7) - (4)}{(7) - (8)}$ | $d_1 = \dfrac{b_1 - b}{b_1 - b_2}$ | 0.7717 | ← |
| (12) | $\dfrac{(4) - (8)}{(7) - (8)}$ | $d_2 = \dfrac{b - b_2}{b_1 - b_2}$ | 0.2283 | ← |
| (13) | Table IV† | $\lambda_{11}$ | 0.9125 | 0.4318 |
| (14) | Table IV† | $\lambda_{12}$ | 0.8899 | 0.4632 |
| (15) | Table IV† | $\lambda_{21}$ | 0.9192 | 0.4408 |
| (16) | Table IV† | $\lambda_{22}$ | 0.8945 | 0.4691 |
| (17) | (10)(13) + (9)(15) | $c_2\lambda_{11} + c_1\lambda_{21}$ | 0.91778 | 0.43370 |
| (18) | (10)(14) + (9)(16) | $c_2\lambda_{12} + c_1\lambda_{22}$ | 0.89353 | 0.46445 |
| (19) | (12)(17) + (11)(18) | $\lambda(ab)$ $= d_2(c_2\lambda_{11} + c_1\lambda_{21})$ $+ d_1(c_2\lambda_{12} + c_1\lambda_{22})$ | 0.8991 | 0.4574 |
| (20) | Sheet 2, row 14, last column† | $t_0$ | 12.47223 | −12.47223 |
| (21). | Sheet 2, row 20, last column | $\sqrt{1 + \dfrac{t_0^2}{2f}}$ | 1.91886 | ← |
| (22) | (20) − (19) × (21) | $t_0 - \lambda\sqrt{1 + \dfrac{t_0^2}{2f}}$ | 10.74699 | −13.34992 |
| (23) | (22) if $\epsilon = L_p$ −(22) if $\epsilon = 1 - L_p$ | $\delta$ | 10.74699 | 13.34992 |
| (24) | Sheet 1, row 5 | $\sqrt{N}$ | 5.47723 | ← |
| (25) | (23) ÷ (24) | $U = \dfrac{\delta}{\sqrt{N}}$ | 1.96212 | 2.43735 |
| (26) | Table 1.5 | $p$ | 0.0249 | 0.0074 |

* The interchange of $a_1$ and $a_2$ in column (4) is made for consistency with the interpolation scheme of sheet 3.

† Use the panel of Table IV that corresponds with the value of $\epsilon$ given in row (2). Note that the portion of Table IV where $t$ is positive is used for column (4) and the portion where $t$ is negative is used for column (5).

If only $N$ and $u_\bullet$ are given and no preliminary calculations are available from work sheets for determining $u_\bullet$ by the exact method, then sheet 4 must be supplemented by a sheet consisting of rows (14) through (24) of sheet 2, with $t$ taken as $\sqrt{N} u_\bullet$ and $f$ taken as $N - 1$.

## 11.5.  Construction of Table 1.2

*Problem.*   *Given: $p_1$, $p_2$, $\alpha$, and $\beta$*
*Required: $N$ and $k$*

A single-sampling plan based on variables is defined by a sample size $N$ and a value $k$, the multiple of the standard deviation in the criterion $\bar{x} + ks < U$ for accepting the lot, where $U$ is the value above which the characteristic being inspected is regarded as defective.   If no restrictions are imposed on sample size, a plan can be constructed for which the operating-characteristic curve will pass through two specified points.   In practice, the first point often corresponds with a lot quality $p_1$ (AQL) for which it is desirable that the probability of rejection should be only $\alpha$ (producer's risk), and the second point corresponds with a lot quality $p_2$ (LTPD) for which it is desirable that the probability of acceptance should be only $\beta$ (consumer's risk).

Consider the family of all OC curves that pass through the point $(p_1, 1 - \alpha)$ but which have differing values of $N$.   If $N$ is small these curves decline gradually as $p$ increases from $p_1$, and if $N$ is large they decline sharply.   There will be one OC curve among the family that passes closer than any other curve to the point $(p_2, \beta)$.   Since $N$ takes only integral values, there usually will be no curve that passes exactly through $(p_2, \beta)$.   A curve could, however, be made to pass exactly through $(p_2, \beta)$ if it were allowed to deviate from $(p_1, 1 - \alpha)$.   Table 1.2 has been constructed in such a way that the OC curves generally pass not exactly through either $(p_1, 1 - \alpha)$ or $(p_2, \beta)$ but very close to both.   The distance of the OC curve from the intended two points has been measured horizontally, that is, by the difference between the actual and intended qualities for which the probabilities of acceptance are $1 - \alpha$ and $\beta$.

The values of $N$ and $k$ in Table 1.2 were obtained in three stages: (1) obtaining a first approximation to $N$ and $k$; (2) determining the degree to which the preliminary values of $N$ and $k$ satisfy the requirement that the OC curve pass through $(p_1, 1 - \alpha)$ and $(p_2, \beta)$; and (3) adjusting $N$ or $k$ to improve the approximation.

*Step* 1.   From $p_1$, $p_2$, $\alpha$, and $\beta$, determine $K_1$, $K_2$, $K_\alpha$, and $K_\beta$, the normal deviates exceeded with probabilities $p_1$, $p_2$, $\alpha$, and $\beta$, respectively, as given by Table 1.1.

*Step* 2.   Calculate the first approximation to $k$ from

$$k = \frac{K_\alpha K_2 + K_\beta K_1}{K_\alpha + K_\beta}$$

*Step* 3.   Calculate an approximate sample size $N$ from

$$N = \frac{k^2 + 2}{2}\left(\frac{K_\alpha + K_\beta}{K_1 - K_2}\right)^2$$

*Step* 4.    By the method given in Sec. 11.4 for determining the quality $p$ corresponding to a specified probability of acceptance $L_p$, determine $p_{1-\alpha}$ and $p_\beta$ where

$p_{1-\alpha}$ = the quality for which the probability of acceptance is $1 - \alpha$

$p_\beta$ = the quality for which the probability of acceptance is $\beta$

If $p_{1-\alpha}$ and $p_\beta$ are sufficiently close to $p_1$ and $p_2$, respectively, then the values of $k$ and $N$ given by steps 2 and 3 define a satisfactory sampling plan.

*Step* 5.    If the discrepancy between $p_{1-\alpha}$ and $p_1$ or that between $p_\beta$ and $p_2$ is too large, determine a better plan by adjusting either $N$ or $k$ according to the following rules:

   *a.* If $p_{1-\alpha}$ is greater than $p_1$ and $p_\beta$ is less than $p_2$, take the next lower integer for $N$ and the same $k$, and repeat step 4.

   *b.* If $p_{1-\alpha}$ is less than $p_1$ and $p_\beta$ is greater than $p_2$, take the next higher integer for $N$ and the same $k$, and repeat step 4.

   *c.* If $p_{1-\alpha}$ is less than $p_1$ and $p_\beta$ is less than $p_2$, use the same $N$ and a smaller value of $k$, and repeat step 4.

   *d.* If $p_{1-\alpha}$ is greater than $p_1$ and $p_\beta$ is greater than $p_2$, use the same $N$ and a larger value of $k$, and repeat step 4.

These rules can be applied regardless of the method used for finding the first approximation to $k$ and $N$. When these are found by steps 2 and 3, however, rule *d* will usually be applicable.

Calculate the better approximation for $k$ mentioned in (*c*) and (*d*) from

$$k = \frac{\lambda_1 K_2 + \lambda_2 K_1}{\lambda_1 + \lambda_2}$$

where

   $\lambda_1$ = value of $\lambda$ computed in determining $p_{1-\alpha}$ in step 4
   $\lambda_2$ = value of $\lambda$ computed in determining $p_\beta$ in step 4

*Step* 6.    For the required $N$ and $k$, take the pair of values for which $p_{1-\alpha}$ and $p_\beta$ are close enough, or closest, to the given values of $p_1$ and $p_2$, respectively.

Table 1.2, which gives values of $N$ and $k$ for various combinations of $p_1$ and $p_2$ when $\alpha = 0.05$ and $\beta = 0.10$, was constructed by essentially the method described above. For each plan in Table 1.2, except that for $p_1 = 0.05$ and $p_2 = 0.40$, the true values of $p_1$ and $p_2$ agree exactly with the intended values when rounded to the number of decimal places with which the intended $p_1$ and $p_2$ are given in the table; in many cases the agreement is much better. Minor improvements could, no doubt, be made in many cases.

### 11.6. Sequential Analysis of Proportion Defective by Variables*

The fact that

$$z = \bar{x} + ks \tag{230}$$

can be taken as approximately normally distributed with mean and standard deviation approximately

$$Z = \mu + k\sigma \tag{231}$$

and

$$\sigma(N) = \sigma F(N) \tag{232}$$

where

$$F(N) = \sqrt{\frac{1}{N} + \frac{k^2}{2(N-1)}} \tag{233}$$

may be made the basis of a sequential procedure for controlling percent defective by means of variables.

For a lot of acceptable quality, that is, one in which the mean of $z$ is $Z_1$ as defined by (155), the probability density of $z$ is given by

$$P(z|Z_1) = \frac{1}{\sigma(N)\sqrt{2\pi}} e^{-\frac{(z-Z_1)^2}{2\sigma^2(N)}} \tag{234}$$

Similarly, for a lot of unacceptable quality, that is, one in which the mean of $z$ is $Z_2$ as defined by (156), the probability density of $z$ is

$$P(z|Z_2) = \frac{1}{\sigma(N)\sqrt{2\pi}} e^{-\frac{(z-Z_1)^2}{2\sigma^2(N)}} \tag{235}$$

A sequential test is made by computing $\lambda$ after each observation, where

$$\lambda = \log_e \frac{P(z|Z_2)}{P(z|Z_1)} \tag{236}$$

and rejecting the lot if

$$\lambda \geq a \tag{237}$$

accepting it if

$$\lambda \leq -b \tag{238}$$

* See Chap. 17 for a brief account of sequential analysis and for references to the fundamental work on this subject, the theory of which was developed in the Statistical Research Group by A. Wald. An exact treatment of proportion defective problems by sequential analysis of variables involves the ratio of densities of two non-central $t$ distributions. Henry Goldberg has shown that this may be expressed in terms of a confluent hypergeometric function, tables of which were prepared for the Statistical Research Group by the New York Mathematical Tables Project in a form suitable for use in sequential analysis. There is in preparation an additional section for *Sequential Analysis* that will include these tables and an explanation of their use. The present section simply outlines a sequential procedure based on the approximations used elsewhere in this chapter for single sampling.

or taking another observation if

$$-b < \lambda < a \tag{239}$$

where

$$a = \log_e \frac{1-\beta}{\alpha} \tag{240}$$

$$b = \log_e \frac{1-\alpha}{\beta} \tag{241}$$

$\alpha$ being the probability of rejecting a lot of quality $Z_1$ and $\beta$ the probability of accepting a lot of quality $Z_2$. Each calculation of $\lambda$ is based on all observations.

To obtain $\lambda$, we take the logarithm of (234) and subtract it from the logarithm of (235), which yields

$$\lambda = \frac{1}{2[\sigma(N)]^2}[(z-Z_1)^2 - (z-Z_2)^2] = \frac{Z_2 - Z_1}{[\sigma(N)]^2}\left(z - \frac{Z_1 + Z_2}{2}\right) \tag{242}$$

Substituting for $Z_1$ and $Z_2$ from (155) and (156) and for $\sigma(N)$ from (232), and rearranging, gives

$$\lambda = \frac{K_1 - K_2}{\sigma[F(N)]^2}\left[z - U - \sigma\left(k - \frac{K_1 + K_2}{2}\right)\right] \tag{243}$$

which may be rewritten as

$$z = U + \sigma\left[\frac{\lambda[F(N)]^2}{K_1 - K_2} - \frac{K_1 + K_2}{2} + k\right] \tag{244}$$

This will be independent of $\sigma$ if

$$k = \frac{K_1 + K_2}{2} - \frac{\lambda[F(N)]^2}{K_1 - K_2} \tag{245}$$

On substituting for $F(N)$ from (233), there results a quadratic in $k$ for which the solution is

$$k = \frac{\sqrt{1 - AB} - 1}{A} \tag{246}$$

where

$$A = \frac{\lambda}{(N-1)(K_1 - K_2)} \tag{247}$$

and

$$B = \frac{2\lambda}{N(K_1 - K_2)} - (K_1 + K_2) \tag{248}$$

For each value of $N$, two values of $k$ are obtained from (246), the first, $k_R$, by replacing $\lambda$ in (247) and (248) by $a$, and the second, $k_A$, by

replacing $\lambda$ by $-b$, where $a$ and $b$ are given by (240) and (241).* Then after each observation both

$$z_R = \bar{x} + k_R s \tag{249}$$

and

$$z_A = \bar{x} + k_A s \tag{250}$$

are computed. If $z_R \geq U$ the lot is rejected. If $z_A \leq U$ the lot is accepted. If neither condition is fulfilled, another item is inspected, new values of $\bar{x}$, $s$, $k_R$, and $k_A$ are computed from all the observations, and new values of $z_R$ and $z_A$ are compared with $U$.

An alternative computational procedure is to compute $k_o$ after each observation, where

$$k_o = \frac{U - \bar{x}}{s} \tag{251}$$

then if $k_o \geq k_A$ the lot is accepted, if $k_o \leq k_R$ the lot is rejected, and if $k_A > k_o > k_R$ another item is inspected.

The values of $k_A$ and $k_R$ depend upon $N$, $p_1$, $p_2$, $\alpha$, and $\beta$. In any one problem all of these but $N$ are fixed, so a table showing $k_A$ and $k_R$ for various values of $N$ can be prepared in advance. If $p_1 = 0.15$, $p_2 = 0.30$, $\alpha = 0.01$ $\beta = 0.02$, and $N = 95$, the values of $k_R$ and $k_A$ are 0.6651 and 0.8930, respectively.† For a single-sampling plan, the corresponding value of $k$ was found in Sec. 4.2.2 to be 0.7645. For the sequential plan, if no decision had been reached from the first 94 observations, rejection would occur if for the first 95 observations $\bar{x} + 0.6651s \geq U$ and acceptance would occur if $\bar{x} + 0.8930s \leq U$; if neither condition were fulfilled, another observation would be made. From the data in Sec. 7.2.2, where $\bar{x} = 30.232$, $s = 5.000$, and $U = 35$, we find $z_R = 33.56$, which, being less than 35, does not call for rejection; and $z_A = 34.70$, which, being less than 35, calls for acceptance.

## 12. NOTATION

The arrangement is alphabetical, the sequence of listing under each letter being as follows: first Roman, then Greek letters; within these

---

* Tables of $a$ and $b$ in terms of $\alpha$ and $\beta$ are given in Statistical Research Group, *Sequential Analysis*, Tables 4.2, 5.1, 6.1, and 7.31. These give $a$ and $b$ to 3 decimal places for values of $\alpha$ and $\beta$ of 0.001, 0.01 to 0.05 by steps of 0.01, 0.05 to 0.20 by steps of 0.05, and 0.20 to 0.40 by steps of 0.10.

† From the unpublished tables referred to in the footnote at the beginning of Sec. 11.6, which are based on the likelihood ratio computed from two non-central $t$ distributions, the corresponding values of $k_R$ and $k_A$ are found to be 0.6630 and 0.8893. For $k_R = 0.6651$ and $k_A = 0.8930$, the true values of $\alpha$ and $\beta$ are, respectively, 0.0110 and 0.0177, which differ by 9.6 and 11.7 percent from the intended values, 0.01 and 0.02. Thus, these two OC points lie below (to the left of) those intended, whereas with single sampling the true points generally lie above (to the right of) those intended (see Table 1.8).

groups, first capital, then small letters; within these groups, first those
without, then those with, primes, bars, or carets; within these groups, first
those without subscripts, next those with literal subscripts (following
the same sequence as for letters), and finally those with numerical
subscripts (in numerical order). Certain symbols, notably $a$ and $b$, are
used in a variety of different meanings that occur only locally; in such
cases the listing is in order of first appearance. Section 11.4 employs the
notation of Johnson and Welch, which differs from (and in some respects,
for example, $U$ and $L$, contradicts) that used elsewhere in the chapter.

| Symbol | See | Meaning |
|---|---|---|
| $A$ | p. 37 | An adjustment constant subtracted from each observation |
| $A$ | p. 61 | Critical value of $z$ for accepting or rejecting a lot |
| $A$ | p. 84 equations (246), (247) | A function of $\lambda$, $N$, $K_1$, and $K_2$ that appears in formulas for the sequential procedure |
| AQL | pp. 11, 45 | Acceptable quality level (see $p_1$) |
| $a$ | pp. 19, 20 equations (21), (22) | Function of $N$ and $K_\alpha$ that appears in the formula for $k$ when $N$ is fixed |
| $a$ | p. 35 equations (55), (56) | Coefficient of quadratic term in condition, quadratic in $X$, for acceptance in terms of $Y$ |
| $a$ | p. 58 equation (118) | Multiple of $\sigma$ in formula for mean of distribution of $s$ |
| $a$ | p. 61 equations (136), (137) | Function of $K_1$, $K_2$, $K_\alpha$, and $K_\beta$ that appears in the formula for $N$ |
| $a$ | pp. 83, 84 equation (240) | A critical value of $\lambda$ in a sequential test on the basis of which the lot is rejected or another observation is taken |
| $\alpha$ | p. 16 | Producer's risk. Probability that a lot of acceptable quality ($p_1$ or AQL) will be rejected |
| $B$ | p. 84 equations (246), (248) | A function of $\lambda$, $N$, $K_1$, and $K_2$ that appears in formulas for the sequential procedure |
| $b$ | pp. 19, 20 equations (21), (23) | Function of $N$, $K_1$, and $K_\alpha$ that appears in the formula for $k$ when $N$ is fixed |
| $b$ | p. 35 equations (55), (57) | Negative of coefficient of linear term in condition, quadratic in $X$, for acceptance in terms of $Y$ |
| $b$ | p. 58 equation (119) | Multiple of $\sigma/\sqrt{2(N-1)}$ in the formula for standard deviation of distribution of $s$ |
| $b$ | p. 61 equations (136), (138) | Function of $k$ and $a$ that appears in the formula for $N$ |
| $b$ | pp. 83, 84 equation (241) | A critical value of $\lambda$ in a sequential test on the basis of which the lot is accepted or another observation is taken |

| Symbol | See | Meaning |
|---|---|---|
| $b_M$ | pp. 53, 54<br>equations (100), (103) | Value of $b$ obtained from equation (57) by replacing $U$ by $M$ |
| $b_U$ | pp. 53, 54<br>equations (100), (103) | Same as $b$ given by equation (57); subscript $U$ used to distinguish $b_U$ from $b_M$ |
| $b'$ | p. 41<br>equations (78), (79) | Negative of coefficient of linear term in condition, linear in $X$, for acceptance in terms of $Y$ |
| $\beta$ | p. 16 | Consumer's risk. Probability that a lot of unacceptable quality ($p_2$ or LTPD) will be accepted |
| $c$ | p. 35<br>equations (55), (58) | Constant term in condition, quadratic in $X$, for acceptance in terms of $Y$ |
| $c_M$ | pp. 53, 54<br>equation (104) | Value of $c$ obtained from equation (58) by replacing $U$ by $M$ |
| $c_U$ | pp. 53, 54<br>equations (100), (104) | Same as $c$ given by equation (58); subscript $U$ used to distinguish $c_U$ from $c_M$ |
| $c'$ | p. 41<br>equations (78), (80) | Constant term in condition, linear in $X$, for acceptance in terms of $Y$ |
| $\gamma$ | p. 43 | A proportion $(0 \leq \gamma \leq 1)$ denoting a confidence coefficient |
| $\gamma_c$ | p. 44 | Confidence coefficient used in judging whether to continue normal inspection |
| $\gamma_r$ | p. 44 | Confidence coefficient used for judging whether to resume normal inspection |
| $\delta$ | p. 73 | Johnson and Welch notation for a constant appearing in the numerator of non-central $t$; $\delta = \sqrt{N} U$ |
| $\delta(f, t_0, \epsilon)$ | p. 73<br>Sec. 11.4.4 | Johnson and Welch notation for the value of $\delta$ that makes $P(f, \delta, t_0) = \epsilon$ |
| $\epsilon$ | p. 17<br>equation (9) | A probability $(0 \leq \epsilon \leq 1)$; probability that a normal deviate will exceed $K_\epsilon$ |
| $\epsilon$ | p. 73 | Johnson and Welch notation for probability of acceptance |
| $F(N)$ | p. 61 | Any function of $N$ that is independent of $\sigma$ |
| $f$ | p. 73 | Degrees of freedom of $\chi^2$ |
| $h$ | p. 42 | Number of lots from which samples are taken and combined into one large sample for the purpose of computing the process average |
| $K$ | p. 9 | $(U - \mu)/\sigma$. That is, the difference between the upper limit and the population mean, as a multiple of the population standard deviation |
| $K_\alpha$ | p. 17 | Normal deviate exceeded with probability $\alpha$ |
| $K_\beta$ | p. 17 | Normal deviate exceeded with probability $\beta$ |
| $K_\epsilon$ | p. 17 | Normal deviate exceeded with probability $\epsilon$ |

| Symbol | See | Meaning |
|---|---|---|
| $K_0$ | p. 21 | Normal deviate exceeded with probability $p_0$ |
| $K_1$ | p. 17 | Normal deviate exceeded with probability $p_1$ |
| $K_2$ | p. 17 | Normal deviate exceeded with probability $p_2$ |
| $k$ | p. 12 | Coefficient of the sample standard deviation in the acceptance criterion $$\bar{x} + ks \leq U;$$ also, acceptable level of statistic $k_o$ |
| $k_A$ | p. 84 | Acceptance level for $k_o$ in a sequential test |
| $k_o$ | p. 43 | $(U - \bar{x})/s$. That is, the observed difference between $U$ and the sample mean in units of the sample standard deviation; sample estimate of $K$ |
| $k_R$ | p. 84 | Rejection level for $k_o$ in a sequential test |
| $k'$ | p. 48 | $\sqrt{n}$ times the coefficient of the sample standard deviation in the acceptance criterion, $\bar{x} + (k'/\sqrt{n}) s \leq U'$, used when items are eventually assembled in sets of $n$ |
| $k_\gamma$ | p. 47 | Upper confidence limit for $K$, based on a confidence coefficient of $\gamma$ |
| $k_{\gamma_a}, k_{\gamma_a}$ | p. 45 | Lower and upper limits, respectively, between which a computed value of $k_o$ leads under normal inspection to continuing normal inspection, based on a confidence coefficient of $\gamma_t$ |
| $k_{\gamma_r}, k_{\gamma_r}$ | p. 45 | Lower and upper limits, respectively, between which a computed value of $k_o$ leads under reduced or tightened inspection to resuming normal inspection, based on a confidence coefficient $\gamma_r$ |
| $L$ | p. 72 | Johnson and Welch notation for the upper limit of a variable, above which an item is defective; corresponds with our $U$ |
| LTPD | pp. 22, 45 | Lot tolerance percent defective (see $p_2$) |
| $L_p$ | p. 15 | Ordinate of an OC curve at abscissa $p$. The probability that a lot of fraction defective $p$ will, if submitted to a given acceptance test, be passed |
| $L_{\mu,\sigma}$ | p. 16 | Value of an OC curve at $\mu, \sigma$. The probability that a lot of mean $\mu$ and standard deviation $\sigma$ will, if submitted to a given acceptance test, be passed |
| $L_1$ | p. 61 | Probability that a lot having mean $Z_1$ will, if submitted to a given acceptance test, be passed |

| Symbol | See | Meaning |
|---|---|---|
| $L_2$ | p. 62 | Probability that a lot having mean $Z_2$ will, if submitted to a given acceptance test, be passed |
| $\lambda$ | p. 83 | Natural logarithm of the ratio of two probability densities, computed after each observation in a sequential test |
| $\lambda = \lambda(f, t_0, \epsilon)$ | p. 74 equations (223), (224) | Johnson and Welch notation for a function of $f$, $t_0$, and $\epsilon$ used in the calculation of $\delta(f, t_0, \epsilon)$ and $t(f, \delta, \epsilon)$ |
| $\lambda_1$ | p. 82 | Value of $\lambda$ computed in determining $p_{1-\alpha}$ |
| $\lambda_2$ | p. 82 | Value of $\lambda$ computed in determining $p_\beta$ |
| $M$ | p. 13 | Minimum value of a variable, below which an item is defective |
| $\mu$ | p. 9 | Population mean of a normally distributed variable |
| $\mu'$ | p. 41 | Best estimate of $\bar{x}$ available before the sample is inspected |
| $N$ | p. 12 | Number of items in a sample |
| $N_a$ | p. 26 | Sample size for single-sampling plan for percent defective based on atributes |
| $N_i$ $(i = 1, \cdots, h)$ | p. 43 | Number of items in the $i$th sample of a group of $h$ samples |
| $N_m$ | p. 27 | Sample size for single-sampling plan for percent defective based on the mean when the standard deviation is known |
| $N_s$ | p. 28 | Sample size for single-sampling plan for percent defective based on the standard deviation when the mean is known |
| $N_t$ | p. 27 | Sample size for single-sampling plan for percent defective based on the mean and standard deviation when both are unknown but estimated from sample |
| $N'$ | p. 48 | Number of sets of $n$ items in a sample |
| $n$ | p. 48 | Number of items constituting a set which is judged defective or not on the basis of the average for the set |
| $n$ | p. 50 | Number of variables upon which the quality of an item depends |
| OC | Sec. 3.4 Sec. 6 | Operating characteristics. The probability that a given acceptance test will pass lots of various qualities if submitted |
| PA | p. 42 | Process average percent defective |
| $P(f, \delta, t_0)$ | p. 73 | Johnson and Welch notation for the probability that $t$ exceeds a given level $t_0$ for given $f$ and $\delta$ |
| $P(z\|Z_1)$ | p. 83 equation (234) | Probability density of $z$ for lot in which the mean of $z$ is $Z_1$ |
| $P(z\|Z_2)$ | p. 83 equation (235) | Probability density of $z$ for lot in which the mean of $z$ is $Z_2$ |

| Symbol | See | Meaning |
|---|---|---|
| $p$ | p. 15 | Fraction defective (quality) of a lot |
| $p_{1-\alpha}$ | p. 82 | The quality for which the probability of acceptance is $1 - \alpha$ with a given sampling plan |
| $p_\beta$ | p. 82 | The quality for which the probability of acceptance is $\beta$ with a given sampling plan |
| $p_0$ | pp. 20, 54 | Indifference quality, or lot fraction defective for which acceptance and rejection are equally probable |
| $p_1$ | p. 16 | Lot fraction defective for which it is desirable that the probability of rejection be at most $\alpha$; (see AQL) |
| $p_2$ | p. 16 | Lot fraction defective for which it is desirable that the probability of acceptance be at most $\beta$; (see LTPD) |
| $p'$ | pp. 30, 33 | Lot fraction defective for which the probability of acceptance ($L_{p'}$) is $1 - L_p$ |
| $p_0', p_0''$ | pp. 55, 64 | Two parts ($p_0' \leq p_0''$) into which $p_0$ is arbitrarily divided, that is, $p_0' + p_0'' = p_0$ |
| $p$ | p. 43 | Point estimate, corresponding with a confidence coefficient of zero, of the process average percent defective. Probability that a normal deviate will exceed $k_o$ |
| $p_\gamma, p_\gamma$ | p. 43 | Lower and upper confidence limits, respectively, for fraction defective where $\gamma$ is the confidence coefficient |
| $S$ | pp. 30, 32 | $\sqrt{1/N + k^2/[2(N-1)]}$.  Quantity used in computations for OC curve |
| $s$ | p. 12 equations (5), (6) | Sample estimate of the standard deviation of a variable |
| $s_i$ $(i = 1, \cdots, h)$ | p. 43 | Sample estimate of the standard deviation of a variable for the $i$th sample of a group of $h$ samples |
| $s^2$ | p. 12 equations (5), (6) | Sample estimate of the variance (squared standard deviation) of a variable |
| $\Sigma$ | p. 12 equations (4), (5) | Summation over a sample.  The variable following $\Sigma$ is to be summed over the sample |
| $\sigma$ | p. 9 | Population standard deviation of a normally distributed variable |
| $\sigma(N)$ | p. 61 equation (139) | Population standard deviation of the normally distributed variable $z$, a function of the sample size $N$ |
| | p. 72 | A variable having the non-central $t$ distribution; in this chapter, used to designate the variable $\sqrt{N}u = (\sqrt{N}/s)(L - \bar{x})$ |
| $t(f, \delta, \epsilon)$ | p. 73 Sec. 11.4.5 | Johnson and Welch notation for that value of $t_0$ for which the probability $P(f, \delta, t_0)$ takes a specified value $\epsilon$ |

| Symbol | See | Meaning |
|---|---|---|
| $t_0$ | p. 73 | Johnson and Welch notation for a given level of non-central $t$ |
| $t_1, t_2,$ | pp. 75, 76 | Successive approximations obtained in the iterative method for calculating $t(f,\delta,\epsilon)$ |
| $U$ | p. 9 | Upper limit of a variable, above which an item is defective |
| $U$ | p. 72 | Johnson and Welch notation for standardized normal deviate exceeded with probability $p$; corresponds with our $K$ |
| $U_1$ | p. 76 | Johnson and Welch notation for the normal deviate exceeded with probability $p_1$ |
| $U'$ | p. 48 | Upper limit of the mean of a set of items, above which the set is defective |
| $U_\gamma$ | p. 47 equation (91) | Tolerance limit on $U$ for a confidence coefficient $\gamma$. Minimum estimate of $U$ sufficient to give a confidence $\gamma$ that the fraction defective will not exceed a fixed $p_1$ |
| $u$ | p. 72 | Johnson and Welch notation for sample statistic $(L - \bar{x})/s$, that is, the observed difference between $L$ and the sample mean in units of the sample standard deviation: corresponds with our $k_o$ |
| $u_0$ | p. 72 | Johnson and Welch notation for critical value of statistic $u$; corresponds with our $k$ |
| $w$ | p. 73 | Johnson and Welch notation for quantity in denominator of noncentral $t$ denoting a variable distributed, independently of $z$, as $\chi^2/f$ with $f$ degrees of freedom; corresponds with our $(s/\sigma)^2$ |
| $X$ | p. 34 | $\Sigma x$, the sum of the observations |
| $\bar{X}$ | p. 67 equation (183) | Value of $X$ at which minimum value $\check{Y}_R$ of $Y_R$ occurs |
| $x_i$ or $x$ $(i = 1, \cdots, N)$ | p. 12 | A sample observation. A measurement on a variable quality characteristic |
| $\bar{x}$ | p. 12 equation (4) | Sample mean of a variable $x$ |
| $\bar{x}_i$ $(i = 1, \cdots, h)$ | p. 43 | Sample mean of a variable $x$ for the $i$th sample of a group of $h$ samples |
| $\bar{x}'$ | p. 56 equation (113) | Abscissa of the point having the same ordinate as the point $(\bar{x},s)$ on the boundary of the acceptance region for a two-sided criterion |
| $\chi^2$ | p. 73 | A variable having the chi-square distribution |
| $Y$ | p. 34 | $\Sigma x^2$, the sum of the squares of the observations |

| Symbol | See | Meaning |
|---|---|---|
| $Y_E$ | p. 36 equation (70) | Boundary between error and acceptance regions when values of $X$ are plotted horizontally and values of $Y$ vertically; minimum value of $Y$ that can be obtained for any given value of $X$ |
| $Y_R$ | p. 36 | Boundary between rejection and acceptance regions when values of $X$ are plotted horizontally and values of $Y$ vertically; parabola defined by equation (55), using the equality sign |
| $Y'$ | p. 37 equation (71) | Sum of squares of adjusted observations, that is, observations from which a constant has been subtracted |
| $Y'_E$ | p. 37 equation (74) | Boundary between the error and acceptance regions when values of $X$ are plotted horizontally and values of $Y'$ vertically |
| $Y'_R$ | p. 37 equation (73) | Boundary between the rejection and acceptance regions when values of $X$ are plotted horizontally and values of $Y'$ vertically |
| $\check{Y}_R$ | p. 67 equation (184) | Minimum value of the curve $Y_R$; a value of $Y$ below which the lot is acceptable without regard to $X$ |
| $\check{Y}'_R$ | p. 69 equation (195) | Minimum value of the curve $Y'_R$; a value of $Y'$ below which the lot is acceptable without regard to $X$ |
| $y$ | p. 74 equation (225) | Johnson and Welch notation for a function of $t_0$ and $f$, one of the arguments against which $\lambda$ is tabled |
| $y'$ | p. 74 equation (226) | Johnson and Welch notation for a function of $t_0$ and $f$, one of the arguments against which $\lambda$ is tabled |
| $Z$ | p. 61 | Population mean of the normally distributed variable $z$ |
| $Z_1$ | p. 61 | Value for $Z$ for which it is required that the probability of rejection be $\alpha$ |
| $Z_2$ | p. 61 | Value of $Z$ for which it is required that the probability of acceptance be $\beta$ |
| $z$ | p. 61 | A normally distributed variable |
| $z$ | p. 73 | Johnson and Welch notation for a variable distributed normally about zero with unit standard deviation |
| $z_A$ | p. 85 | $\bar{x} + k_A s$ |
| $z_R$ | p. 85 | $\bar{x} + k_R s$ |

## 13. REFERENCES

American Society for Testing Materials: *Manual of Presentation of Data*, Philadelphia, 1941.

American Standards Association: *Control Chart Method of Controlling Quality during Production*, New York, 1942, 41 pp.

COMRIE, L. J., editor: *Barlow's Tables of Squares, Cubes, Square Roots, Cube Roots, and Reciprocals of All Integral Numbers up to 12,500*, 4th ed., E. & F. N. Spon, Ltd., London, and Chemical Publishing Company of New York, Inc., New York, 1941, 258 + xii pp.

CRAIG, CECIL C.: "Note on the Distribution of Non-central $t$ with an Application," *Annals of Mathematical Statistics*, Vol. 12 (1941), pp. 224–228.

DODGE, HAROLD F., and ROMIG, HARRY G.: *Sampling Inspection Tables: Single and Double Sampling*, John Wiley & Sons, Inc., New York, 1944, 106 + vi pp.

FISHER, RONALD A., and YATES, FRANK: *Statistical Tables for Biological, Agricultural and Medical Research*, 2d ed., Oliver and Boyd, Ltd., London and Edinburgh, 1943, 98 + viii pp.

JENNETT, W. J., and WELCH, B. L.: "The Control of Proportion Defective As Judged by a Single Quality Characteristic Varying on a Continuous Scale," *Supplement to the Journal of the Royal Statistical Society*, Vol. 6 (1939), pp. 80–88.

JOHNSON, N. L., and WELCH, B. L.: "Applications of the Non-central $t$-Distribution," *Biometrika*, Vol. 31 (1940), pp. 362–389.

KELLEY, TRUMAN LEE: *The Kelley Statistical Tables*, The Macmillan Company, New York, 1938, 136 pp.

Mathematical Tables Project, Work Projects Administration for the City of New York, Federal Works Agency: *Tables of Probability Functions*, Vol. 2, National Bureau of Standards, Washington, D.C., 1942, 344 + xxi pp.

PEACH, PAUL: *An Introduction to Industrial Statistics and Quality Control*, 2d ed., Edwards & Broughton Company, Raleigh, N. C., 1947, 236 + xvi pp.

PRZYBOROWSKI, J., and WILEŃSKI, H.: "Statistical Principles of Routine Work in Testing Clover Seed for Dodder," *Biometrika*, Vol. 27 (1935), pp. 273–292.

ROMIG, HARRY G.: *Allowable Average in Sampling Inspection*, privately printed doctoral thesis, Columbia University, 1939, 58 pp.

SHEWHART, WALTER A.: *Economic Control of Quality of Manufactured Product*, D. Van Nostrand Company, Inc., New York, 1931, 501 + xiv pp.

———, with the editorial assistance of DEMING, W. EDWARDS: *Statistical Method from the Viewpoint of Quality Control*, Graduate School, U.S. Department of Agriculture, Washington, D. C., 1939, 155 + ix pp.

Statistical Research Group, Columbia University: *Sampling Inspection: Principles, Procedures, and Tables for Single, Double, and Sequential Sampling in Acceptance Inspection and Quality Control*, McGraw-Hill Book Company, Inc., New York, 1948.

———: *Sequential Analysis of Statistical Data: Applications*, rev. ed., Columbia University Press, New York, 1945.

# CHAPTER 2

# TOLERANCE LIMITS FOR NORMAL DISTRIBUTIONS

This chapter was prepared by
Albert H. Bowker, Stanford University

# CHAPTER 2

# TOLERANCE LIMITS FOR NORMAL DISTRIBUTIONS

## 1. NATURE OF TOLERANCE LIMITS

### 1.1. Definition of Tolerance Limits

A common method of specifying the quality of manufactured product is to set a range, the bounds of which are called "tolerance limits," such that a certain percentage of the product made under commercial conditions may be expected to have a quality falling between the limits. For example, tolerance limits may be placed on dimensions to indicate the extent of interchangeability of parts. Tolerance limits are also used in other fields, for example medicine and psychology, to indicate the range of normal variation in various characteristics. In general, tolerance limits are limits within which lie a certain percentage of the individuals of a statistical universe.

The satisfactory application of tolerance limits requires that assignable causes of variability be detected and eliminated so that the remaining variability may be considered random. In other words, prediction from past data of the proportion of future observations that will be contained within fixed limits assumes that the successive values of the quality characteristic in question may be considered randomly chosen from a statistical universe.*

### 1.2. Class of Tolerance Limits Considered

The tolerance limits considered in this chapter are symmetrical about the mean, and they are based on the assumption that the underlying statistical universe may be described by a normal distribution. When the true mean and standard deviation $\mu$ and $\sigma$ of this normal distribution are known, tolerance limits are formed by adding to and subtracting from the mean a certain multiple $K$ of the standard deviation; the appropriate multiple is easily obtained from tables of the normal distribution and depends only upon what proportion of the universe is to be included within the limits.† Values of $K$ corresponding to various proportions

---

* A detailed discussion of the nature, use, and practical basis for tolerance limits may be found in Shewhart, *Statistical Method*, Chap. II, pp. 50–79. In fact, this is the original source on which subsequent work on tolerance limits has been based. (For full citations, see Sec. 4.)

† See Chap. 1, Sec. 2, for further discussion of this point.

*included* within the tolerance range may be found by entering Table 1.1 of Chap. 1 with $\epsilon$ equal to half the proportion *excluded*. For example, 90 percent of the. population is included in the interval $\mu - 1.645\sigma$ to $\mu + 1.645\sigma$, since 90 percent inclusion corresponds with 10 percent exclusion, or $\epsilon = 0.05$.

In most practical situations the true values $\mu$ and $\sigma$ are not known, and tolerance limits must be estimated from the mean $\bar{X}$ and the standard deviation $s$ of a sample of $N$ observations. While the limits $\mu \pm 1.645\sigma$ will include 90 percent of the universe, clearly the same assertion cannot be made of the limits $\bar{X} \pm 1.645s$. The proportion included between the limits $\bar{X} \pm Ks$ depends, for example, on whether the sample mean is unusually large, reasonably close to the population mean, or unusually small. The sample standard deviation is also subject to chance fluctuations that affect the proportion included between the limits. In fact, if a value of $K$ is fixed, no two samples from a given universe will lead to the same tolerance limits, since both $\bar{X}$ and $s$ will vary from sample to sample.

It is clear then that the proportion of the universe included between the limits $\bar{X} \pm Ks$ is a random variable, and that it is impossible to determine $K$ so that the limits will always contain a specified proportion $P$ of the universe. It is possible, however, to determine $K$ in such a way that in a large series of samples from normal universes a certain proportion $\gamma$ of the intervals $\bar{X} \pm Ks$ will include $P$ or more of the universe; $\gamma$ is called "the confidence coefficient," since it is a measure of the confidence with which we may assert that a given tolerance range includes at least $P$ of the universe.

Intuitively, it seems reasonable that the values of $K$ for use with sample means and standard deviations will be somewhat larger than the values used if the true mean and standard deviation are known. In fact, by taking $K$ very large the probability that the interval $\bar{X} \pm Ks$ will contain at least $P$ of the universe may be made as close to 1 as is desired. Since the length of the tolerance interval is $2Ks$, the smaller $K$ is, the more meaningful and the less trivial a tolerance range is. The position is, therefore, one of being able to state broad generalities with little risk of error, or to make more precise statements at greater risk. The problem is to find the smallest factor $K$ that may be associated with a specified confidence coefficient $\gamma$, proportion $P$, and sample size $N$. Table 2.1 presents such factors for various combinations of $\gamma$, $P$, and $N$. Before discussing the use of Table 2.1, it is desirable to contrast the tolerance limits it yields with other kinds of limits.

### 1.3. Other Kinds of Limits

**1.3.1. One-sided Tolerance Limits.**—Instead of specifying a two-sided interval that will contain a certain percentage of the product, it

will frequently be appropriate to specify a one-sided limit such that a fixed percentage of the product will have a quality greater than (or a quality less than, as the case may be) this limit. Problems of this sort may be solved by the methods of Chap. 1, Sec. 9.2.

**1.3.2. Multivariate Tolerance Limits.**—Certain methods of setting tolerance limits have been generalized by Wald and by Paulson to the problem of setting simultaneous limits on several quality characteristics of a single item. In this connection, see also Chap. 3.

**1.3.3. Tolerance Limits for Non-normal Distributions.**—Wald has considered the problem of setting tolerance limits on the basis of large samples for distributions of any known form, normal or otherwise. Application of these tolerance limits to normal distributions gives unsatisfactory results; their applicability (as contrasted with their theoretical correctness when $N$ approaches infinity) to other forms of distribution has not been determined.

**1.3.4. Distribution-free Tolerance Limits.**—Sometimes it is desirable to set tolerance limits without making any assumption about the form, normal or otherwise, of the underlying distribution. Problems of this sort have been discussed by Wilks and by Wald. The problem of deciding whether to use tolerance limits based on the normal distribution, those based on some other distribution, or those which involve no assumption about the form of the distribution is one which, though of a kind common in applied statistics, has no satisfactory solution. Tolerance limits based on the normal distribution are substantially shorter for a fixed sample size than those based on no assumption about the distribution; but they may be irrelevant if the underlying distribution is too far from normal. The decision about the adequacy of the normality assumption may be based upon an analysis of measurements of a large number of items; it may be supported by theoretical considerations, for example, when the quality characteristic is known to be determined by many chance causes whose effects are essentially additive; or it may rest on a combination of theoretical and empirical evidence. It should be noted that if there is a stable universe from which the measurements are random observations, limits of the form $\bar{X} \pm Ks$ will, in a fraction $\gamma$ of samples, include a certain proportion $P'$ or more of the universe, where $P'$ differs from the intended $P$ by an amount depending upon the departure of the true distribution from normality.*

**1.3.5. Confidence Limits.**—Tolerance limits should not be confused with confidence limits (or with fiducial limits) for the mean or other parameters of the distribution. Confidence limits for the mean, which also are of the form $\bar{X} \pm Ks$, are computed in such a way that in a fraction $\gamma$ of the samples from which they are computed they will include the mean

---

* For a fuller and more precise statement of this point, see Chap. 1, Sec. 9.4, including the quotation from Jennett and Welch.

of the distribution.    Tolerance limits, in contrast, are computed in such a way that in a fraction $\gamma$ of the samples from which they are computed they will include at·least a fraction $P$ of the items in the distribution.    Thus, as the sample size increases, the value of $K$ for a confidence interval diminishes toward zero, whereas the value of $K$ for a tolerance interval diminishes toward the normal deviate including a proportion $P$ of the universe.

## 2. TABLE OF FACTORS FOR TOLERANCE LIMITS

### 2.1. Description

Table 2.1 shows 3,880 values of the multiple $K$ of the sample standard deviation $s$ to be added to and subtracted from the sample mean $\overline{X}$, for various values of the confidence coefficient $\gamma$, the proportion $P$ of the distribution to be included within the tolerance limits and the sample size $N$.    Four values of $\gamma$ are given: 0.75, 0.90, 0.95, and 0.99; five values of $P$: 0.75, 0.90, 0.95, 0.99, and 0.999; and 194 values of $N$: from 2 to 102 by steps of 1, from 102 to 180 by steps of 2, from 180 to 300 by steps of 5, from 300 to 400 by steps of 10, from 400 to 750 by steps of 25, from 750 to 1,000 by steps of 50, and infinity.

### 2.2. Use of the Table : Example

In connection with fire-control problems for a certain projectile it is desirable to know, with a confidence coefficient of 0.99, limits within which the muzzle velocities of 90 percent of the future rounds will lie.    The required calculations are performed on the observed velocities as follows: First, add all the observations, calling the sum $\Sigma X$.    Then add all the squares of the observations, calling the sum $\Sigma X^2$.    These two sums and $N$, the number of observations, are the only information needed from the sample; in the present example they are

$$\Sigma X = 291,200 \text{ ft. per sec.} \tag{1}$$
$$\Sigma X^2 = 393,114,400 \tag{2}$$
$$N = 216 \tag{3}$$

The mean and standard deviation are computed as follows:

$$\overline{X} = \frac{\Sigma X}{N} = 1,348.15 \text{ ft. per sec.} \tag{4}$$

$$s = \sqrt{\frac{N\Sigma X^2 - (\Sigma X)^2}{N(N-1)}} = 49.82 \text{ ft. per sec.} \tag{5}$$

Entering Table 2.1 with $P = 0.90$ and $\gamma = 0.99$, we find for $N = 215$, $K = 1.856$, and for $N = 220$, $K = 1.854$; by linear interpolation we find that for $N = 216$, $K = 1.856$.*    The required tolerance interval, in feet

* Linear interpolation with respect to $N$ will not result in an error of more than 1 in the third decimal place.

per second, is then $1,348.15 \pm 1.856 \times 49.82$; so we may assert with 99 percent confidence that 90 per cent of all the future observations will lie between 1,255.7 and 1,440.6 ft. per sec.

## 2.3. Cases outside the Range of the Table

**2.3.1. Formula for Large Values of $N$.**—For values of $N$ beyond the range of the table, the following formula may be used:*

$$K = K_\infty \left( 1 + \frac{x_\gamma}{\sqrt{2N}} + \frac{5x_\gamma^2 + 10}{12N} \right) \tag{6}$$

where $K_\infty$ (the value of $K$ we would use if the true $\mu$ and $\sigma$ were known) is defined by

$$\frac{1}{\sqrt{2}\pi} \int_{K_\infty}^{K} e^{-\frac{t^2}{2}} dt = P \tag{7}$$

and $x_\gamma$ is defined by

$$\frac{1}{\sqrt{2}\pi} \int_{x_\gamma}^{\infty} e^{-\frac{t^2}{2}} dt = 1 - \gamma \tag{8}$$

Values of $K_\infty$ are listed in Table 2.1 under the appropriate $P$ columns and are the same for all $\gamma$; they may also be determined easily from Table 1.1 in Chap. 1. Values of $x_\gamma$ may be found from any table of the areas of the normal curve, such as Table 1.1. For the values of $\gamma$ used in Table 2.1, however, Table 2.2 gives the values of the functions of $x$ required within the parentheses of formula (6).

Unless $N$ is too small, formula (6) also gives results of satisfactory accuracy for values of $P$ and $\gamma$ not included in Table 2.1. When $N \geq 160$, $\gamma \leq 0.99$, and $P \leq 0.999$, formula (6) gives values of $K$ agreeing within 1 unit in the third significant figure with those in Table 2.1.

**2.3.2. Example.**—In studying another characteristic of the projectile referred to in Sec. 2.2, the following sample values were obtained:

$$N = 1,449 \tag{9}$$
$$\bar{X} = 670.40 \tag{10}$$
$$s = 37.45 \tag{11}$$

Again assuming $\gamma = 0.99$ and $P = 0.90$, we have, from Table 2.2 and formula (6),

$$K = 1.645 \left( 1 + \frac{1.6450}{\sqrt{1449}} + \frac{3.0883}{1449} \right) = 1.720 \tag{12}$$

The upper and lower tolerance limits are, therefore, 734.81 and 605.99, respectively.

* This formula was developed in the Statistical Research Group by Albert H. Bowker; see his paper cited in Sec. 4.

## TABLE 2.1

### TOLERANCE FACTORS FOR NORMAL DISTRIBUTIONS

Factors $K$ such that the probability is $\gamma$ that at least a proportion $P$ of the distribution will be included between $\overline{X} \pm Ks$, where $\overline{X}$ and $s$ are estimates of the mean and the standard deviation computed from a sample of $N$.

| P / N | $\gamma = 0.75$ | | | | | $\gamma = 0.90$ | | | | | $\gamma = 0.95$ | | | | | $\gamma = 0.99$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0.75 | 0.90 | 0.95 | 0.99 | 0.999 | 0.75 | 0.90 | 0.95 | 0.99 | 0.999 | 0.75 | 0.90 | 0.95 | 0.99 | 0.999 | 0.75 | 0.90 | 0.95 | 0.99 | 0.999 |
| 2 | 4.498 | 6.301 | 7.414 | 9.531 | 11.920 | 11.407 | 15.978 | 18.800 | 24.167 | 30.227 | 22.858 | 32.019 | 37.674 | 48.430 | 60.573 | 114.363 | 160.193 | 188.491 | 242.300 | 303.054 |
| 3 | 2.501 | 3.538 | 4.187 | 5.431 | 6.844 | 4.132 | 5.847 | 6.919 | 8.974 | 11.309 | 5.922 | 8.380 | 9.916 | 12.861 | 16.208 | 13.378 | 18.930 | 22.401 | 29.055 | 36.616 |
| 4 | 2.035 | 2.892 | 3.431 | 4.471 | 5.657 | 2.932 | 4.166 | 4.943 | 6.440 | 8.149 | 3.779 | 5.369 | 6.370 | 8.299 | 10.502 | 6.614 | 9.398 | 11.150 | 14.527 | 18.383 |
| 5 | 1.825 | 2.599 | 3.088 | 4.033 | 5.117 | 2.454 | 3.494 | 4.152 | 5.423 | 6.879 | 3.002 | 4.275 | 5.079 | 6.634 | 8.415 | 4.643 | 6.612 | 7.855 | 10.260 | 13.015 |
| 6 | 1.704 | 2.429 | 2.889 | 3.779 | 4.802 | 2.196 | 3.131 | 3.723 | 4.870 | 6.188 | 2.604 | 3.712 | 4.414 | 5.775 | 7.337 | 3.743 | 5.337 | 6.345 | 8.301 | 10.548 |
| 7 | 1.624 | 2.318 | 2.757 | 3.611 | 4.593 | 2.034 | 2.902 | 3.452 | 4.521 | 5.750 | 2.361 | 3.369 | 4.007 | 5.248 | 6.676 | 3.233 | 4.613 | 5.488 | 7.187 | 9.142 |
| 8 | 1.568 | 2.238 | 2.663 | 3.491 | 4.444 | 1.921 | 2.743 | 3.264 | 4.278 | 5.446 | 2.197 | 3.136 | 3.732 | 4.891 | 6.226 | 2.905 | 4.147 | 4.936 | 6.468 | 8.234 |
| 9 | 1.525 | 2.178 | 2.593 | 3.400 | 4.330 | 1.839 | 2.626 | 3.125 | 4.098 | 5.220 | 2.078 | 2.967 | 3.532 | 4.631 | 5.899 | 2.677 | 3.822 | 4.550 | 5.966 | 7.600 |
| 10 | 1.492 | 2.131 | 2.537 | 3.328 | 4.241 | 1.775 | 2.535 | 3.018 | 3.959 | 5.046 | 1.987 | 2.839 | 3.379 | 4.433 | 5.649 | 2.508 | 3.582 | 4.265 | 5.594 | 7.129 |
| 11 | 1.465 | 2.093 | 2.493 | 3.271 | 4.169 | 1.724 | 2.463 | 2.933 | 3.849 | 4.906 | 1.916 | 2.737 | 3.259 | 4.277 | 5.452 | 2.378 | 3.397 | 4.045 | 5.308 | 6.766 |
| 12 | 1.443 | 2.062 | 2.456 | 3.223 | 4.110 | 1.683 | 2.404 | 2.863 | 3.758 | 4.792 | 1.858 | 2.655 | 3.162 | 4.150 | 5.291 | 2.274 | 3.250 | 3.870 | 5.079 | 6.477 |
| 13 | 1.425 | 2.036 | 2.424 | 3.183 | 4.059 | 1.648 | 2.355 | 2.805 | 3.682 | 4.697 | 1.810 | 2.587 | 3.081 | 4.044 | 5.158 | 2.190 | 3.130 | 3.727 | 4.893 | 6.240 |
| 14 | 1.409 | 2.013 | 2.398 | 3.148 | 4.016 | 1.619 | 2.314 | 2.756 | 3.618 | 4.615 | 1.770 | 2.529 | 3.012 | 3.955 | 5.045 | 2.120 | 3.029 | 3.608 | 4.737 | 6.043 |
| 15 | 1.395 | 1.994 | 2.375 | 3.118 | 3.979 | 1.594 | 2.278 | 2.713 | 3.562 | 4.545 | 1.735 | 2.480 | 2.954 | 3.878 | 4.949 | 2.060 | 2.945 | 3.507 | 4.605 | 5.876 |
| 16 | 1.383 | 1.977 | 2.355 | 3.092 | 3.946 | 1.572 | 2.246 | 2.676 | 3.514 | 4.484 | 1.705 | 2.437 | 2.903 | 3.812 | 4.865 | 2.009 | 2.872 | 3.421 | 4.492 | 5.732 |
| 17 | 1.372 | 1.962 | 2.337 | 3.069 | 3.917 | 1.552 | 2.219 | 2.643 | 3.471 | 4.430 | 1.679 | 2.400 | 2.858 | 3.754 | 4.791 | 1.965 | 2.808 | 3.345 | 4.393 | 5.607 |
| 18 | 1.363 | 1.948 | 2.321 | 3.048 | 3.891 | 1.535 | 2.194 | 2.614 | 3.433 | 4.382 | 1.655 | 2.366 | 2.819 | 3.702 | 4.725 | 1.926 | 2.753 | 3.279 | 4.307 | 5.497 |
| 19 | 1.355 | 1.936 | 2.307 | 3.030 | 3.867 | 1.520 | 2.172 | 2.588 | 3.399 | 4.339 | 1.635 | 2.337 | 2.784 | 3.656 | 4.667 | 1.891 | 2.703 | 3.221 | 4.230 | 5.399 |
| 20 | 1.347 | 1.925 | 2.294 | 3.013 | 3.846 | 1.506 | 2.152 | 2.564 | 3.368 | 4.300 | 1.616 | 2.310 | 2.752 | 3.615 | 4.614 | 1.860 | 2.659 | 3.168 | 4.161 | 5.312 |
| 21 | 1.340 | 1.915 | 2.282 | 2.998 | 3.827 | 1.493 | 2.135 | 2.543 | 3.340 | 4.264 | 1.599 | 2.286 | 2.723 | 3.577 | 4.567 | 1.833 | 2.620 | 3.121 | 4.100 | 5.234 |
| 22 | 1.334 | 1.906 | 2.271 | 2.984 | 3.809 | 1.482 | 2.118 | 2.524 | 3.315 | 4.232 | 1.584 | 2.264 | 2.697 | 3.543 | 4.523 | 1.808 | 2.584 | 3.078 | 4.044 | 5.163 |
| 23 | 1.328 | 1.898 | 2.261 | 2.971 | 3.793 | 1.471 | 2.103 | 2.506 | 3.292 | 4.203 | 1.570 | 2.244 | 2.673 | 3.512 | 4.484 | 1.785 | 2.551 | 3.040 | 3.993 | 5.098 |
| 24 | 1.322 | 1.891 | 2.252 | 2.959 | 3.778 | 1.462 | 2.089 | 2.489 | 3.270 | 4.176 | 1.557 | 2.225 | 2.651 | 3.483 | 4.447 | 1.764 | 2.522 | 3.004 | 3.947 | 5.039 |
| 25 | 1.317 | 1.883 | 2.244 | 2.948 | 3.764 | 1.453 | 2.077 | 2.474 | 3.251 | 4.151 | 1.545 | 2.208 | 2.631 | 3.457 | 4.413 | 1.745 | 2.494 | 2.972 | 3.904 | 4.985 |
| 26 | 1.313 | 1.877 | 2.236 | 2.938 | 3.751 | 1.444 | 2.065 | 2.460 | 3.232 | 4.127 | 1.534 | 2.193 | 2.612 | 3.432 | 4.382 | 1.727 | 2.469 | 2.941 | 3.865 | 4.935 |
| 27 | 1.309 | 1.871 | 2.229 | 2.929 | 3.740 | 1.437 | 2.054 | 2.447 | 3.215 | 4.106 | 1.523 | 2.178 | 2.595 | 3.409 | 4.353 | 1.711 | 2.446 | 2.914 | 3.828 | 4.888 |

TABLE 2.1     TOLERANCE LIMITS     103

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 28 | 4.845 | 3.794 | 2.888 | 2.424 | 1.695 | 4.326 | 3.388 | 2.579 | 2.164 | 1.514 | 4.085 | 3.199 | 2.435 | 2.044 | 1.430 | 3.728 | 2.920 | 2.222 | 1.865 | 1.305 |
| 29 | 4.805 | 3.763 | 2.864 | 2.404 | 1.681 | 4.301 | 3.368 | 2.564 | 2.152 | 1.505 | 4.066 | 3.184 | 2.424 | 2.034 | 1.423 | 3.718 | 2.911 | 2.216 | 1.860 | 1.301 |
| 30 | 4.768 | 3.733 | 2.841 | 2.385 | 1.668 | 4.278 | 3.350 | 2.549 | 2.140 | 1.497 | 4.049 | 3.170 | 2.413 | 2.025 | 1.417 | 3.708 | 2.904 | 2.210 | 1.855 | 1.297 |
| 31 | 4.733 | 3.706 | 2.820 | 2.367 | 1.656 | 4.256 | 3.332 | 2.536 | 2.129 | 1.489 | 4.032 | 3.157 | 2.403 | 2.017 | 1.411 | 3.699 | 2.896 | 2.204 | 1.850 | 1.294 |
| 32 | 4.699 | 3.680 | 2.801 | 2.351 | 1.644 | 4.235 | 3.316 | 2.524 | 2.118 | 1.481 | 4.016 | 3.145 | 2.393 | 2.009 | 1.405 | 3.690 | 2.890 | 2.199 | 1.846 | 1.291 |
| 33 | 4.668 | 3.655 | 2.782 | 2.335 | 1.633 | 4.215 | 3.300 | 2.512 | 2.108 | 1.475 | 4.001 | 3.133 | 2.385 | 2.001 | 1.400 | 3.682 | 2.883 | 2.194 | 1.842 | 1.288 |
| 34 | 4.639 | 3.632 | 2.764 | 2.320 | 1.623 | 4.197 | 3.286 | 2.501 | 2.099 | 1.468 | 3.987 | 3.122 | 2.376 | 1.994 | 1.395 | 3.674 | 2.877 | 2.189 | 1.838 | 1.285 |
| 35 | 4.611 | 3.611 | 2.748 | 2.306 | 1.613 | 4.179 | 3.272 | 2.490 | 2.090 | 1.462 | 3.974 | 3.112 | 2.368 | 1.988 | 1.390 | 3.667 | 2.871 | 2.185 | 1.834 | 1.283 |
| 36 | 4.585 | 3.590 | 2.732 | 2.293 | 1.604 | 4.161 | 3.258 | 2.479 | 2.081 | 1.455 | 3.961 | 3.102 | 2.361 | 1.981 | 1.386 | 3.660 | 2.866 | 2.181 | 1.830 | 1.280 |
| 37 | 4.560 | 3.571 | 2.717 | 2.281 | 1.595 | 4.146 | 3.246 | 2.470 | 2.073 | 1.450 | 3.949 | 3.092 | 2.353 | 1.975 | 1.381 | 3.653 | 2.860 | 2.177 | 1.827 | 1.278 |
| 38 | 4.537 | 3.552 | 2.703 | 2.269 | 1.587 | 4.134 | 3.237 | 2.464 | 2.068 | 1.446 | 3.938 | 3.083 | 2.346 | 1.969 | 1.377 | 3.647 | 2.855 | 2.173 | 1.824 | 1.275 |
| 39 | 4.514 | 3.534 | 2.690 | 2.257 | 1.579 | 4.120 | 3.226 | 2.455 | 2.060 | 1.441 | 3.927 | 3.075 | 2.340 | 1.964 | 1.374 | 3.641 | 2.850 | 2.169 | 1.821 | 1.273 |
| 40 | 4.493 | 3.518 | 2.677 | 2.247 | 1.571 | 4.104 | 3.213 | 2.445 | 2.052 | 1.435 | 3.917 | 3.066 | 2.334 | 1.959 | 1.370 | 3.635 | 2.846 | 2.166 | 1.818 | 1.271 |
| 41 | 4.472 | 3.502 | 2.665 | 2.236 | 1.564 | 4.090 | 3.202 | 2.437 | 2.045 | 1.430 | 3.907 | 3.059 | 2.328 | 1.954 | 1.366 | 3.629 | 2.841 | 2.162 | 1.815 | 1.269 |
| 42 | 4.453 | 3.486 | 2.653 | 2.227 | 1.557 | 4.077 | 3.192 | 2.429 | 2.039 | 1.426 | 3.897 | 3.051 | 2.322 | 1.949 | 1.363 | 3.624 | 2.837 | 2.159 | 1.812 | 1.267 |
| 43 | 4.434 | 3.472 | 2.642 | 2.217 | 1.551 | 4.065 | 3.183 | 2.422 | 2.033 | 1.422 | 3.888 | 3.044 | 2.316 | 1.944 | 1.360 | 3.619 | 2.833 | 2.156 | 1.810 | 1.266 |
| 44 | 4.416 | 3.458 | 2.631 | 2.208 | 1.545 | 4.053 | 3.173 | 2.413 | 2.027 | 1.418 | 3.879 | 3.037 | 2.311 | 1.940 | 1.357 | 3.614 | 2.829 | 2.153 | 1.807 | 1.264 |
| 45 | 4.399 | 3.444 | 2.621 | 2.200 | 1.539 | 4.042 | 3.165 | 2.406 | 2.021 | 1.414 | 3.871 | 3.030 | 2.306 | 1.935 | 1.354 | 3.609 | 2.826 | 2.150 | 1.805 | 1.262 |
| 46 | 4.383 | 3.431 | 2.611 | 2.192 | 1.533 | 4.031 | 3.156 | 2.402 | 2.016 | 1.410 | 3.863 | 3.024 | 2.301 | 1.931 | 1.351 | 3.605 | 2.822 | 2.148 | 1.802 | 1.261 |
| 47 | 4.367 | 3.419 | 2.602 | 2.184 | 1.527 | 4.021 | 3.148 | 2.396 | 2.011 | 1.406 | 3.855 | 3.018 | 2.297 | 1.927 | 1.348 | 3.600 | 2.819 | 2.145 | 1.800 | 1.259 |
| 48 | 4.352 | 3.407 | 2.593 | 2.176 | 1.522 | 4.011 | 3.140 | 2.390 | 2.006 | 1.402 | 3.847 | 3.012 | 2.292 | 1.924 | 1.345 | 3.596 | 2.815 | 2.143 | 1.798 | 1.258 |
| 49 | 4.337 | 3.396 | 2.584 | 2.169 | 1.517 | 4.002 | 3.133 | 2.384 | 2.001 | 1.399 | 3.840 | 3.006 | 2.288 | 1.920 | 1.343 | 3.592 | 2.812 | 2.140 | 1.796 | 1.256 |
| 50 | 4.323 | 3.385 | 2.576 | 2.162 | 1.512 | 3.993 | 3.126 | 2.379 | 1.996 | 1.396 | 3.833 | 3.001 | 2.284 | 1.916 | 1.340 | 3.588 | 2.809 | 2.138 | 1.794 | 1.255 |
| 51 | 4.310 | 3.374 | 2.568 | 2.155 | 1.507 | 3.984 | 3.116 | 2.373 | 1.992 | 1.393 | 3.826 | 2.995 | 2.279 | 1.913 | 1.338 | 3.584 | 2.806 | 2.135 | 1.792 | 1.253 |
| 52 | 4.297 | 3.364 | 2.560 | 2.148 | 1.503 | 3.975 | 3.112 | 2.368 | 1.988 | 1.390 | 3.820 | 2.990 | 2.276 | 1.910 | 1.336 | 3.581 | 2.803 | 2.133 | 1.790 | 1.252 |
| 53 | 4.284 | 3.354 | 2.552 | 2.142 | 1.498 | 3.967 | 3.106 | 2.363 | 1.984 | 1.387 | 3.813 | 2.985 | 2.272 | 1.907 | 1.334 | 3.577 | 2.801 | 2.131 | 1.789 | 1.251 |
| 54 | 4.272 | 3.344 | 2.545 | 2.136 | 1.494 | 3.959 | 3.100 | 2.359 | 1.980 | 1.384 | 3.807 | 2.981 | 2.268 | 1.904 | 1.331 | 3.574 | 2.798 | 2.129 | 1.787 | 1.250 |
| 55 | 4.260 | 3.335 | 2.538 | 2.130 | 1.490 | 3.951 | 3.094 | 2.354 | 1.976 | 1.382 | 3.801 | 2.976 | 2.265 | 1.901 | 1.329 | 3.571 | 2.795 | 2.127 | 1.785 | 1.249 |
| 56 | 4.249 | 3.326 | 2.531 | 2.124 | 1.486 | 3.944 | 3.088 | 2.350 | 1.972 | 1.379 | 3.796 | 2.972 | 2.261 | 1.898 | 1.327 | 3.567 | 2.793 | 2.125 | 1.784 | 1.247 |
| 57 | 4.238 | 3.318 | 2.524 | 2.119 | 1.482 | 3.937 | 3.082 | 2.345 | 1.968 | 1.377 | 3.790 | 2.967 | 2.258 | 1.895 | 1.325 | 3.564 | 2.790 | 2.123 | 1.782 | 1.246 |
| 58 | 4.227 | 3.309 | 2.518 | 2.113 | 1.478 | 3.930 | 3.076 | 2.341 | 1.965 | 1.374 | 3.785 | 2.963 | 2.255 | 1.892 | 1.323 | 3.561 | 2.788 | 2.122 | 1.781 | 1.245 |
| 59 | 4.216 | 3.301 | 2.512 | 2.108 | 1.474 | 3.923 | 3.071 | 2.337 | 1.961 | 1.372 | 3.779 | 2.959 | 2.252 | 1.890 | 1.322 | 3.558 | 2.786 | 2.120 | 1.779 | 1.244 |
| 60 | 4.206 | 3.293 | 2.506 | 2.103 | 1.471 | 3.916 | 3.066 | 2.333 | 1.958 | 1.369 | 3.774 | 2.955 | 2.248 | 1.887 | 1.320 | 3.556 | 2.784 | 2.118 | 1.778 | 1.243 |
| 61 | 4.196 | 3.285 | 2.500 | 2.098 | 1.467 | 3.909 | 3.061 | 2.329 | 1.955 | 1.367 | 3.769 | 2.951 | 2.245 | 1.885 | 1.318 | 3.553 | 2.781 | 2.117 | 1.776 | 1.242 |
| 62 | 4.187 | 3.278 | 2.494 | 2.093 | 1.464 | 3.903 | 3.056 | 2.325 | 1.951 | 1.365 | 3.765 | 2.947 | 2.243 | 1.882 | 1.316 | 3.550 | 2.779 | 2.115 | 1.775 | 1.241 |
| 63 | 4.178 | 3.271 | 2.489 | 2.089 | 1.461 | 3.897 | 3.051 | 2.322 | 1.948 | 1.363 | 3.760 | 2.944 | 2.240 | 1.880 | 1.315 | 3.548 | 2.777 | 2.113 | 1.774 | 1.240 |
| 64 | 4.169 | 3.264 | 2.483 | 2.084 | 1.458 | 3.891 | 3.046 | 2.318 | 1.945 | 1.361 | 3.755 | 2.940 | 2.237 | 1.878 | 1.313 | 3.545 | 2.775 | 2.112 | 1.772 | 1.240 |

TABLE 2.1.—(Continued)

| N \ P | γ = 0.75 |  |  |  |  | γ = 0.90 |  |  |  |  | γ = 0.95 |  |  |  |  | γ = 0.99 |  |  |  |  |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
|  | 0.75 | 0.90 | 0.95 | 0.99 | 0.999 | 0.75 | 0.90 | 0.95 | 0.99 | 0.999 | 0.75 | 0.90 | 0.95 | 0.99 | 0.999 | 0.75 | 0.90 | 0.95 | 0.99 | 0.999 |
| 65 | 1.239 | 1.771 | 2.110 | 2.773 | 3.543 | 1.312 | 1.875 | 2.235 | 2.937 | 3.751 | 1.359 | 1.943 | 2.315 | 3.042 | 3.886 | 1.455 | 2.080 | 2.478 | 3.257 | 4.160 |
| 66 | 1.238 | 1.770 | 2.109 | 2.771 | 3.540 | 1.310 | 1.873 | 2.232 | 2.933 | 3.747 | 1.357 | 1.940 | 2.311 | 3.037 | 3.880 | 1.452 | 2.076 | 2.473 | 3.250 | 4.152 |
| 67 | 1.237 | 1.769 | 2.108 | 2.770 | 3.538 | 1.309 | 1.871 | 2.229 | 2.930 | 3.742 | 1.355 | 1.937 | 2.308 | 3.033 | 3.874 | 1.449 | 2.071 | 2.468 | 3.244 | 4.143 |
| 68 | 1.236 | 1.768 | 2.106 | 2.768 | 3.536 | 1.307 | 1.869 | 2.227 | 2.927 | 3.738 | 1.353 | 1.934 | 2.305 | 3.029 | 3.869 | 1.446 | 2.067 | 2.463 | 3.237 | 4.135 |
| 69 | 1.235 | 1.766 | 2.105 | 2.766 | 3.533 | 1.306 | 1.867 | 2.225 | 2.923 | 3.734 | 1.351 | 1.932 | 2.302 | 3.025 | 3.864 | 1.443 | 2.063 | 2.459 | 3.231 | 4.127 |
| 70 | 1.235 | 1.765 | 2.104 | 2.764 | 3.531 | 1.304 | 1.865 | 2.222 | 2.920 | 3.730 | 1.349 | 1.929 | 2.299 | 3.021 | 3.859 | 1.440 | 2.060 | 2.454 | 3.225 | 4.120 |
| 71 | 1.234 | 1.764 | 2.102 | 2.763 | 3.529 | 1.303 | 1.863 | 2.220 | 2.917 | 3.727 | 1.347 | 1.927 | 2.296 | 3.017 | 3.854 | 1.438 | 2.056 | 2.450 | 3.219 | 4.112 |
| 72 | 1.233 | 1.763 | 2.101 | 2.761 | 3.527 | 1.302 | 1.861 | 2.218 | 2.915 | 3.723 | 1.346 | 1.924 | 2.293 | 3.013 | 3.849 | 1.435 | 2.052 | 2.445 | 3.214 | 4.105 |
| 73 | 1.233 | 1.762 | 2.100 | 2.760 | 3.525 | 1.300 | 1.859 | 2.216 | 2.912 | 3.719 | 1.344 | 1.922 | 2.290 | 3.009 | 3.844 | 1.433 | 2.049 | 2.441 | 3.208 | 4.098 |
| 74 | 1.232 | 1.761 | 2.099 | 2.758 | 3.523 | 1.299 | 1.858 | 2.214 | 2.909 | 3.716 | 1.343 | 1.920 | 2.287 | 3.006 | 3.840 | 1.430 | 2.045 | 2.437 | 3.203 | 4.091 |
| 75 | 1.231 | 1.760 | 2.098 | 2.757 | 3.521 | 1.298 | 1.856 | 2.211 | 2.906 | 3.712 | 1.341 | 1.917 | 2.285 | 3.002 | 3.835 | 1.428 | 2.042 | 2.433 | 3.197 | 4.084 |
| 76 | 1.230 | 1.759 | 2.096 | 2.755 | 3.519 | 1.297 | 1.854 | 2.209 | 2.904 | 3.709 | 1.339 | 1.915 | 2.282 | 2.999 | 3.831 | 1.426 | 2.039 | 2.429 | 3.192 | 4.078 |
| 77 | 1.230 | 1.758 | 2.095 | 2.754 | 3.517 | 1.296 | 1.853 | 2.207 | 2.901 | 3.706 | 1.338 | 1.913 | 2.279 | 2.996 | 3.826 | 1.423 | 2.035 | 2.425 | 3.187 | 4.071 |
| 78 | 1.229 | 1.758 | 2.094 | 2.752 | 3.516 | 1.295 | 1.851 | 2.206 | 2.898 | 3.702 | 1.336 | 1.911 | 2.277 | 2.992 | 3.822 | 1.421 | 2.032 | 2.421 | 3.182 | 4.065 |
| 79 | 1.229 | 1.757 | 2.093 | 2.751 | 3.514 | 1.293 | 1.849 | 2.204 | 2.896 | 3.699 | 1.335 | 1.909 | 2.274 | 2.989 | 3.818 | 1.419 | 2.029 | 2.418 | 3.177 | 4.059 |
| 80 | 1.228 | 1.756 | 2.092 | 2.749 | 3.512 | 1.292 | 1.848 | 2.202 | 2.894 | 3.696 | 1.334 | 1.907 | 2.272 | 2.986 | 3.814 | 1.417 | 2.026 | 2.414 | 3.173 | 4.053 |
| 81 | 1.227 | 1.755 | 2.091 | 2.748 | 3.510 | 1.291 | 1.846 | 2.200 | 2.891 | 3.693 | 1.332 | 1.905 | 2.270 | 2.983 | 3.810 | 1.415 | 2.023 | 2.411 | 3.168 | 4.047 |
| 82 | 1.227 | 1.754 | 2.090 | 2.747 | 3.509 | 1.290 | 1.845 | 2.198 | 2.889 | 3.690 | 1.331 | 1.903 | 2.267 | 2.980 | 3.806 | 1.413 | 2.020 | 2.407 | 3.163 | 4.041 |
| 83 | 1.226 | 1.753 | 2.089 | 2.746 | 3.507 | 1.289 | 1.843 | 2.196 | 2.887 | 3.687 | 1.329 | 1.901 | 2.265 | 2.977 | 3.803 | 1.411 | 2.017 | 2.404 | 3.159 | 4.035 |
| 84 | 1.226 | 1.753 | 2.088 | 2.744 | 3.506 | 1.288 | 1.842 | 2.195 | 2.884 | 3.684 | 1.328 | 1.899 | 2.263 | 2.974 | 3.799 | 1.409 | 2.014 | 2.400 | 3.155 | 4.030 |
| 85 | 1.225 | 1.752 | 2.087 | 2.743 | 3.504 | 1.287 | 1.841 | 2.193 | 2.882 | 3.682 | 1.327 | 1.897 | 2.261 | 2.971 | 3.795 | 1.407 | 2.012 | 2.397 | 3.150 | 4.024 |
| 86 | 1.225 | 1.751 | 2.086 | 2.742 | 3.503 | 1.286 | 1.839 | 2.191 | 2.880 | 3.679 | 1.326 | 1.896 | 2.259 | 2.968 | 3.792 | 1.405 | 2.009 | 2.394 | 3.146 | 4.019 |
| 87 | 1.224 | 1.750 | 2.086 | 2.741 | 3.501 | 1.285 | 1.838 | 2.190 | 2.878 | 3.676 | 1.324 | 1.894 | 2.257 | 2.966 | 3.788 | 1.403 | 2.007 | 2.391 | 3.142 | 4.014 |
| 88 | 1.224 | 1.749 | 2.085 | 2.740 | 3.500 | 1.284 | 1.837 | 2.188 | 2.876 | 3.674 | 1.323 | 1.892 | 2.255 | 2.963 | 3.785 | 1.402 | 2.004 | 2.388 | 3.138 | 4.009 |
| 89 | 1.223 | 1.749 | 2.084 | 2.738 | 3.498 | 1.284 | 1.835 | 2.187 | 2.874 | 3.671 | 1.322 | 1.890 | 2.253 | 2.960 | 3.781 | 1.400 | 2.001 | 2.385 | 3.134 | 4.004 |
| 90 | 1.223 | 1.748 | 2.083 | 2.737 | 3.497 | 1.283 | 1.834 | 2.185 | 2.872 | 3.669 | 1.321 | 1.889 | 2.251 | 2.958 | 3.778 | 1.398 | 1.999 | 2.382 | 3.130 | 3.999 |
| 91 | 1.222 | 1.747 | 2.082 | 2.736 | 3.495 | 1.282 | 1.833 | 2.184 | 2.870 | 3.666 | 1.320 | 1.887 | 2.249 | 2.955 | 3.775 | 1.396 | 1.997 | 2.379 | 3.127 | 3.994 |
| 92 | 1.222 | 1.747 | 2.081 | 2.735 | 3.494 | 1.281 | 1.832 | 2.182 | 2.868 | 3.664 | 1.319 | 1.886 | 2.247 | 2.953 | 3.772 | 1.395 | 1.994 | 2.376 | 3.123 | 3.989 |
| 93 | 1.221 | 1.746 | 2.081 | 2.734 | 3.493 | 1.280 | 1.830 | 2.181 | 2.866 | 3.661 | 1.318 | 1.884 | 2.245 | 2.950 | 3.769 | 1.393 | 1.992 | 2.373 | 3.119 | 3.985 |

TABLE 2.1　　　　　TOLERANCE LIMITS　　　　　105

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 94 | 3.980 | 3.116 | 2.371 | 1.990 | 1.392 | 3.766 | 2.948 | 2.243 | 1.882 | 1.317 | 3.659 | 2.864 | 2.180 | 1.829 | 1.279 | 3.491 | 2.733 | 2.080 | 1.745 | 1.221 |
| 95 | 3.976 | 3.112 | 2.368 | 1.987 | 1.390 | 3.763 | 2.945 | 2.241 | 1.881 | 1.315 | 3.657 | 2.863 | 2.178 | 1.828 | 1.278 | 3.490 | 2.732 | 2.079 | 1.745 | 1.220 |
| 96 | 3.971 | 3.109 | 2.366 | 1.985 | 1.388 | 3.760 | 2.943 | 2.240 | 1.880 | 1.314 | 3.654 | 2.861 | 2.177 | 1.827 | 1.278 | 3.489 | 2.731 | 2.078 | 1.744 | 1.220 |
| 97 | 3.967 | 3.105 | 2.363 | 1.983 | 1.387 | 3.757 | 2.941 | 2.238 | 1.878 | 1.313 | 3.652 | 2.859 | 2.175 | 1.826 | 1.277 | 3.488 | 2.730 | 2.077 | 1.743 | 1.219 |
| 98 | 3.963 | 3.102 | 2.360 | 1.981 | 1.385 | 3.754 | 2.939 | 2.236 | 1.877 | 1.312 | 3.650 | 2.857 | 2.174 | 1.825 | 1.276 | 3.486 | 2.729 | 2.077 | 1.743 | 1.219 |
| 99 | 3.958 | 3.099 | 2.358 | 1.979 | 1.384 | 3.751 | 2.936 | 2.234 | 1.875 | 1.311 | 3.648 | 2.856 | 2.173 | 1.824 | 1.275 | 3.485 | 2.728 | 2.076 | 1.742 | 1.219 |
| 100 | 3.954 | 3.096 | 2.355 | 1.977 | 1.383 | 3.748 | 2.934 | 2.233 | 1.874 | 1.311 | 3.646 | 2.854 | 2.172 | 1.822 | 1.275 | 3.484 | 2.727 | 2.075 | 1.742 | 1.218 |
| 101 | 3.950 | 3.092 | 2.353 | 1.975 | 1.381 | 3.746 | 2.932 | 2.231 | 1.872 | 1.310 | 3.644 | 2.852 | 2.170 | 1.821 | 1.274 | 3.483 | 2.726 | 2.075 | 1.741 | 1.218 |
| 102 | 3.946 | 3.089 | 2.351 | 1.973 | 1.380 | 3.743 | 2.930 | 2.230 | 1.871 | 1.309 | 3.642 | 2.851 | 2.169 | 1.820 | 1.273 | 3.482 | 2.726 | 2.074 | 1.741 | 1.217 |
| 104 | 3.939 | 3.083 | 2.346 | 1.969 | 1.377 | 3.738 | 2.926 | 2.227 | 1.869 | 1.307 | 3.638 | 2.848 | 2.167 | 1.818 | 1.272 | 3.480 | 2.724 | 2.073 | 1.739 | 1.217 |
| 106 | 3.931 | 3.077 | 2.342 | 1.965 | 1.374 | 3.733 | 2.922 | 2.224 | 1.866 | 1.305 | 3.634 | 2.845 | 2.164 | 1.816 | 1.270 | 3.477 | 2.722 | 2.071 | 1.738 | 1.216 |
| 108 | 3.924 | 3.072 | 2.337 | 1.962 | 1.372 | 3.728 | 2.918 | 2.221 | 1.864 | 1.303 | 3.630 | 2.842 | 2.162 | 1.815 | 1.269 | 3.475 | 2.721 | 2.070 | 1.737 | 1.215 |
| 110 | 3.917 | 3.066 | 2.333 | 1.958 | 1.369 | 3.723 | 2.915 | 2.218 | 1.861 | 1.302 | 3.626 | 2.839 | 2.160 | 1.813 | 1.268 | 3.473 | 2.719 | 2.069 | 1.736 | 1.214 |
| 112 | 3.910 | 3.061 | 2.329 | 1.955 | 1.367 | 3.719 | 2.911 | 2.215 | 1.859 | 1.300 | 3.623 | 2.836 | 2.158 | 1.811 | 1.267 | 3.471 | 2.717 | 2.068 | 1.735 | 1.214 |
| 114 | 3.904 | 3.056 | 2.325 | 1.951 | 1.365 | 3.714 | 2.908 | 2.212 | 1.857 | 1.299 | 3.619 | 2.833 | 2.156 | 1.809 | 1.265 | 3.469 | 2.716 | 2.067 | 1.734 | 1.213 |
| 116 | 3.897 | 3.051 | 2.321 | 1.948 | 1.363 | 3.710 | 2.904 | 2.210 | 1.855 | 1.297 | 3.616 | 2.831 | 2.154 | 1.808 | 1.264 | 3.468 | 2.714 | 2.065 | 1.733 | 1.212 |
| 118 | 3.891 | 3.046 | 2.318 | 1.945 | 1.360 | 3.706 | 2.901 | 2.207 | 1.852 | 1.296 | 3.613 | 2.828 | 2.152 | 1.806 | 1.263 | 3.466 | 2.713 | 2.064 | 1.733 | 1.212 |
| 120 | 3.885 | 3.041 | 2.314 | 1.942 | 1.358 | 3.702 | 2.898 | 2.205 | 1.850 | 1.294 | 3.610 | 2.826 | 2.150 | 1.804 | 1.262 | 3.464 | 2.712 | 2.063 | 1.732 | 1.211 |
| 122 | 3.879 | 3.037 | 2.311 | 1.939 | 1.356 | 3.695 | 2.895 | 2.203 | 1.848 | 1.293 | 3.607 | 2.823 | 2.148 | 1.803 | 1.261 | 3.462 | 2.710 | 2.062 | 1.731 | 1.211 |
| 124 | 3.873 | 3.032 | 2.307 | 1.936 | 1.354 | 3.694 | 2.892 | 2.200 | 1.847 | 1.291 | 3.604 | 2.821 | 2.147 | 1.801 | 1.260 | 3.461 | 2.709 | 2.061 | 1.731 | 1.210 |
| 126 | 3.868 | 3.028 | 2.304 | 1.934 | 1.352 | 3.690 | 2.889 | 2.198 | 1.845 | 1.290 | 3.601 | 2.819 | 2.145 | 1.800 | 1.259 | 3.459 | 2.708 | 2.060 | 1.729 | 1.209 |
| 128 | 3.862 | 3.024 | 2.301 | 1.931 | 1.350 | 3.686 | 2.886 | 2.196 | 1.843 | 1.289 | 3.598 | 2.816 | 2.143 | 1.799 | 1.258 | 3.458 | 2.707 | 2.060 | 1.728 | 1.209 |
| 130 | 3.857 | 3.019 | 2.298 | 1.928 | 1.349 | 3.683 | 2.883 | 2.194 | 1.841 | 1.288 | 3.595 | 2.814 | 2.141 | 1.797 | 1.257 | 3.456 | 2.705 | 2.059 | 1.728 | 1.208 |
| 132 | 3.852 | 3.015 | 2.295 | 1.926 | 1.347 | 3.679 | 2.880 | 2.192 | 1.839 | 1.286 | 3.592 | 2.812 | 2.140 | 1.796 | 1.256 | 3.455 | 2.704 | 2.058 | 1.727 | 1.208 |
| 134 | 3.847 | 3.012 | 2.292 | 1.923 | 1.345 | 3.676 | 2.878 | 2.190 | 1.838 | 1.285 | 3.590 | 2.810 | 2.138 | 1.795 | 1.255 | 3.453 | 2.703 | 2.057 | 1.726 | 1.207 |
| 136 | 3.842 | 3.008 | 2.289 | 1.921 | 1.343 | 3.673 | 2.875 | 2.188 | 1.836 | 1.284 | 3.587 | 2.808 | 2.137 | 1.793 | 1.254 | 3.452 | 2.702 | 2.056 | 1.725 | 1.207 |
| 138 | 3.837 | 3.004 | 2.286 | 1.918 | 1.342 | 3.669 | 2.873 | 2.186 | 1.834 | 1.283 | 3.585 | 2.806 | 2.135 | 1.792 | 1.253 | 3.450 | 2.701 | 2.055 | 1.725 | 1.206 |
| 140 | 3.833 | 3.000 | 2.283 | 1.916 | 1.340 | 3.666 | 2.870 | 2.184 | 1.833 | 1.282 | 3.582 | 2.804 | 2.134 | 1.791 | 1.252 | 3.449 | 2.700 | 2.054 | 1.724 | 1.206 |
| 142 | 3.828 | 2.997 | 2.280 | 1.914 | 1.338 | 3.663 | 2.868 | 2.182 | 1.831 | 1.281 | 3.580 | 2.802 | 2.132 | 1.790 | 1.251 | 3.448 | 2.699 | 2.054 | 1.723 | 1.205 |
| 144 | 3.824 | 2.993 | 2.278 | 1.912 | 1.337 | 3.660 | 2.865 | 2.180 | 1.830 | 1.280 | 3.578 | 2.801 | 2.131 | 1.788 | 1.250 | 3.446 | 2.698 | 2.053 | 1.723 | 1.205 |
| 146 | 3.820 | 2.990 | 2.275 | 1.909 | 1.335 | 3.657 | 2.863 | 2.178 | 1.828 | 1.279 | 3.575 | 2.799 | 2.130 | 1.787 | 1.249 | 3.445 | 2.697 | 2.052 | 1.722 | 1.204 |
| 148 | 3.815 | 2.987 | 2.273 | 1.907 | 1.334 | 3.654 | 2.861 | 2.177 | 1.827 | 1.278 | 3.573 | 2.797 | 2.128 | 1.786 | 1.248 | 3.444 | 2.696 | 2.051 | 1.722 | 1.204 |
| 150 | 3.811 | 2.983 | 2.270 | 1.905 | 1.332 | 3.652 | 2.859 | 2.175 | 1.825 | 1.277 | 3.571 | 2.795 | 2.127 | 1.785 | 1.248 | 3.443 | 2.695 | 2.051 | 1.721 | 1.204 |
| 152 | 3.807 | 2.980 | 2.268 | 1.903 | 1.331 | 3.649 | 2.856 | 2.173 | 1.824 | 1.276 | 3.569 | 2.794 | 2.126 | 1.784 | 1.247 | 3.441 | 2.694 | 2.050 | 1.720 | 1.203 |
| 154 | 3.803 | 2.977 | 2.265 | 1.901 | 1.330 | 3.646 | 2.854 | 2.172 | 1.823 | 1.275 | 3.567 | 2.792 | 2.125 | 1.783 | 1.247 | 3.440 | 2.693 | 2.049 | 1.720 | 1.203 |
| 156 | 3.799 | 2.974 | 2.263 | 1.899 | 1.328 | 3.644 | 2.852 | 2.170 | 1.821 | 1.274 | 3.565 | 2.791 | 2.123 | 1.782 | 1.246 | 3.439 | 2.692 | 2.049 | 1.719 | 1.202 |
| 158 | 3.796 | 2.971 | 2.261 | 1.897 | 1.327 | 3.641 | 2.850 | 2.169 | 1.820 | 1.273 | 3.563 | 2.789 | 2.122 | 1.781 | 1.246 | 3.438 | 2.691 | 2.048 | 1.719 | 1.202 |

TABLE 2.1.—(Continued)

| N | γ = 0.75 | | | | | γ = 0.90 | | | | | γ = 0.95 | | | | | γ = 0.99 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| P | 0.75 | 0.90 | 0.95 | 0.99 | 0.999 | 0.75 | 0.90 | 0.95 | 0.99 | 0.999 | 0.75 | 0.90 | 0.95 | 0.99 | 0.999 | 0.75 | 0.90 | 0.95 | 0.99 | 0.999 |
| 160 | 1.202 | 1.718 | 2.047 | 2.691 | 3.437 | 1.245 | 1.780 | 2.121 | 2.787 | 3.561 | 1.272 | 1.819 | 2.167 | 2.848 | 3.638 | 1.326 | 1.896 | 2.259 | 2.968 | 3.792 |
| 162 | 1.201 | 1.718 | 2.047 | 2.690 | 3.436 | 1.244 | 1.779 | 2.120 | 2.786 | 3.559 | 1.271 | 1.818 | 2.166 | 2.846 | 3.636 | 1.324 | 1.894 | 2.256 | 2.965 | 3.788 |
| 164 | 1.201 | 1.717 | 2.046 | 2.689 | 3.435 | 1.244 | 1.778 | 2.119 | 2.785 | 3.557 | 1.270 | 1.816 | 2.164 | 2.844 | 3.634 | 1.323 | 1.892 | 2.254 | 2.963 | 3.785 |
| 166 | 1.201 | 1.717 | 2.045 | 2.688 | 3.434 | 1.243 | 1.777 | 2.118 | 2.783 | 3.555 | 1.269 | 1.815 | 2.163 | 2.843 | 3.631 | 1.322 | 1.890 | 2.252 | 2.960 | 3.781 |
| 168 | 1.200 | 1.716 | 2.045 | 2.687 | 3.433 | 1.242 | 1.776 | 2.117 | 2.782 | 3.553 | 1.269 | 1.814 | 2.162 | 2.841 | 3.629 | 1.321 | 1.888 | 2.250 | 2.957 | 3.778 |
| 170 | 1.200 | 1.716 | 2.044 | 2.687 | 3.432 | 1.242 | 1.775 | 2.116 | 2.780 | 3.552 | 1.268 | 1.813 | 2.160 | 2.839 | 3.627 | 1.320 | 1.887 | 2.248 | 2.955 | 3.774 |
| 172 | 1.199 | 1.715 | 2.044 | 2.686 | 3.431 | 1.241 | 1.775 | 2.115 | 2.779 | 3.550 | 1.267 | 1.812 | 2.159 | 2.837 | 3.624 | 1.318 | 1.885 | 2.246 | 2.952 | 3.771 |
| 174 | 1.199 | 1.715 | 2.043 | 2.685 | 3.430 | 1.240 | 1.774 | 2.114 | 2.778 | 3.548 | 1.266 | 1.811 | 2.158 | 2.835 | 3.622 | 1.317 | 1.884 | 2.244 | 2.950 | 3.768 |
| 176 | 1.199 | 1.714 | 2.043 | 2.684 | 3.429 | 1.240 | 1.773 | 2.113 | 2.776 | 3.547 | 1.266 | 1.810 | 2.156 | 2.834 | 3.620 | 1.316 | 1.882 | 2.243 | 2.947 | 3.765 |
| 178 | 1.199 | 1.714 | 2.042 | 2.684 | 3.428 | 1.239 | 1.772 | 2.112 | 2.775 | 3.545 | 1.265 | 1.809 | 2.155 | 2.832 | 3.618 | 1.315 | 1.880 | 2.241 | 2.945 | 3.762 |
| 180 | 1.198 | 1.713 | 2.042 | 2.683 | 3.427 | 1.239 | 1.771 | 2.111 | 2.774 | 3.543 | 1.264 | 1.808 | 2.154 | 2.831 | 3.616 | 1.314 | 1.879 | 2.239 | 2.942 | 3.759 |
| 185 | 1.197 | 1.712 | 2.040 | 2.681 | 3.425 | 1.237 | 1.769 | 2.108 | 2.771 | 3.539 | 1.262 | 1.805 | 2.151 | 2.827 | 3.611 | 1.311 | 1.875 | 2.234 | 2.937 | 3.751 |
| 190 | 1.197 | 1.711 | 2.039 | 2.680 | 3.423 | 1.236 | 1.767 | 2.106 | 2.768 | 3.536 | 1.261 | 1.803 | 2.148 | 2.823 | 3.606 | 1.309 | 1.872 | 2.230 | 2.931 | 3.744 |
| 195 | 1.196 | 1.710 | 2.038 | 2.678 | 3.421 | 1.235 | 1.766 | 2.104 | 2.765 | 3.532 | 1.259 | 1.800 | 2.145 | 2.819 | 3.601 | 1.307 | 1.868 | 2.226 | 2.926 | 3.738 |
| 200 | 1.195 | 1.709 | 2.037 | 2.677 | 3.419 | 1.234 | 1.764 | 2.102 | 2.762 | 3.529 | 1.258 | 1.798 | 2.143 | 2.816 | 3.597 | 1.304 | 1.865 | 2.222 | 2.921 | 3.731 |
| 205 | 1.195 | 1.708 | 2.036 | 2.675 | 3.418 | 1.233 | 1.762 | 2.100 | 2.760 | 3.526 | 1.256 | 1.796 | 2.140 | 2.812 | 3.593 | 1.302 | 1.862 | 2.219 | 2.916 | 3.725 |
| 210 | 1.194 | 1.708 | 2.035 | 2.674 | 3.416 | 1.231 | 1.761 | 2.098 | 2.757 | 3.522 | 1.255 | 1.794 | 2.138 | 2.809 | 3.589 | 1.300 | 1.859 | 2.215 | 2.911 | 3.719 |
| 215 | 1.194 | 1.707 | 2.034 | 2.673 | 3.414 | 1.230 | 1.759 | 2.096 | 2.755 | 3.519 | 1.253 | 1.792 | 2.135 | 2.806 | 3.586 | 1.298 | 1.856 | 2.212 | 2.907 | 3.713 |
| 220 | 1.193 | 1.706 | 2.033 | 2.671 | 3.413 | 1.229 | 1.758 | 2.095 | 2.753 | 3.516 | 1.252 | 1.790 | 2.133 | 2.803 | 3.581 | 1.296 | 1.854 | 2.209 | 2.903 | 3.708 |
| 225 | 1.192 | 1.705 | 2.032 | 2.670 | 3.411 | 1.228 | 1.756 | 2.093 | 2.750 | 3.514 | 1.251 | 1.788 | 2.131 | 2.800 | 3.577 | 1.294 | 1.851 | 2.205 | 2.898 | 3.703 |
| 230 | 1.192 | 1.704 | 2.031 | 2.669 | 3.409 | 1.227 | 1.755 | 2.091 | 2.748 | 3.511 | 1.249 | 1.787 | 2.129 | 2.798 | 3.574 | 1.293 | 1.848 | 2.202 | 2.894 | 3.697 |
| 235 | 1.191 | 1.704 | 2.030 | 2.668 | 3.408 | 1.226 | 1.754 | 2.090 | 2.746 | 3.508 | 1.248 | 1.785 | 2.127 | 2.795 | 3.571 | 1.291 | 1.846 | 2.199 | 2.890 | 3.692 |
| 240 | 1.191 | 1.703 | 2.029 | 2.667 | 3.407 | 1.226 | 1.752 | 2.088 | 2.744 | 3.506 | 1.247 | 1.783 | 2.125 | 2.792 | 3.567 | 1.289 | 1.843 | 2.197 | 2.887 | 3.688 |
| 245 | 1.190 | 1.702 | 2.028 | 2.666 | 3.405 | 1.225 | 1.751 | 2.087 | 2.742 | 3.503 | 1.246 | 1.782 | 2.123 | 2.790 | 3.564 | 1.288 | 1.841 | 2.194 | 2.883 | 3.683 |
| 250 | 1.190 | 1.702 | 2.028 | 2.665 | 3.404 | 1.224 | 1.750 | 2.085 | 2.740 | 3.501 | 1.245 | 1.780 | 2.121 | 2.788 | 3.561 | 1.286 | 1.839 | 2.191 | 2.880 | 3.678 |
| 255 | 1.190 | 1.701 | 2.027 | 2.664 | 3.403 | 1.223 | 1.749 | 2.084 | 2.739 | 3.499 | 1.244 | 1.779 | 2.119 | 2.785 | 3.558 | 1.284 | 1.837 | 2.189 | 2.876 | 3.674 |
| 260 | 1.189 | 1.700 | 2.026 | 2.663 | 3.401 | 1.222 | 1.748 | 2.083 | 2.737 | 3.496 | 1.243 | 1.777 | 2.118 | 2.783 | 3.555 | 1.283 | 1.835 | 2.186 | 2.873 | 3.670 |
| 265 | 1.189 | 1.700 | 2.025 | 2.662 | 3.400 | 1.222 | 1.747 | 2.081 | 2.735 | 3.494 | 1.242 | 1.776 | 2.116 | 2.781 | 3.552 | 1.282 | 1.833 | 2.184 | 2.870 | 3.666 |
| 270 | 1.188 | 1.699 | 2.025 | 2.661 | 3.399 | 1.221 | 1.746 | 2.080 | 2.734 | 3.492 | 1.241 | 1.774 | 2.114 | 2.779 | 3.550 | 1.280 | 1.831 | 2.181 | 2.867 | 3.662 |

TABLE 2.1 TOLERANCE LIMITS 107

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 275 | 3.658 | 2.864 | 2.179 | 1.829 | 1.279 | 3.547 | 2.777 | 2.113 | 1.773 | 1.240 | 3.490 | 2.732 | 2.079 | 1.745 | 1.220 | 3.398 | 2.660 | 2.024 | 1.699 | 1.188 |
| 280 | 3.655 | 2.861 | 2.177 | 1.827 | 1.278 | 3.544 | 2.775 | 2.111 | 1.772 | 1.239 | 3.488 | 2.730 | 2.078 | 1.744 | 1.219 | 3.397 | 2.659 | 2.023 | 1.698 | 1.188 |
| 285 | 3.651 | 2.858 | 2.175 | 1.825 | 1.275 | 3.542 | 2.773 | 2.110 | 1.771 | 1.238 | 3.486 | 2.729 | 2.076 | 1.743 | 1.219 | 3.396 | 2.658 | 2.022 | 1.698 | 1.187 |
| 290 | 3.647 | 2.855 | 2.173 | 1.823 | 1.274 | 3.540 | 2.771 | 2.108 | 1.769 | 1.237 | 3.484 | 2.727 | 2.075 | 1.742 | 1.218 | 3.395 | 2.657 | 2.022 | 1.697 | 1.186 |
| 295 | 3.644 | 2.853 | 2.170 | 1.822 | 1.274 | 3.537 | 2.769 | 2.107 | 1.768 | 1.237 | 3.482 | 2.726 | 2.074 | 1.741 | 1.217 | 3.394 | 2.656 | 2.021 | 1.696 | 1.186 |
| 300 | 3.641 | 2.850 | 2.169 | 1.820 | 1.273 | 3.535 | 2.767 | 2.106 | 1.767 | 1.236 | 3.481 | 2.725 | 2.073 | 1.740 | 1.217 | 3.393 | 2.656 | 2.021 | 1.696 | 1.186 |
| 310 | 3.634 | 2.845 | 2.165 | 1.817 | 1.271 | 3.531 | 2.764 | 2.103 | 1.765 | 1.234 | 3.477 | 2.722 | 2.071 | 1.738 | 1.216 | 3.391 | 2.655 | 2.020 | 1.695 | 1.185 |
| 320 | 3.628 | 2.840 | 2.161 | 1.814 | 1.268 | 3.526 | 2.760 | 2.100 | 1.763 | 1.233 | 3.474 | 2.719 | 2.069 | 1.737 | 1.215 | 3.389 | 2.653 | 2.019 | 1.694 | 1.185 |
| 330 | 3.623 | 2.836 | 2.158 | 1.811 | 1.266 | 3.522 | 2.757 | 2.098 | 1.761 | 1.231 | 3.471 | 2.717 | 2.067 | 1.735 | 1.213 | 3.388 | 2.652 | 2.018 | 1.693 | 1.184 |
| 340 | 3.617 | 2.831 | 2.154 | 1.808 | 1.265 | 3.519 | 2.754 | 2.096 | 1.759 | 1.230 | 3.468 | 2.715 | 2.066 | 1.734 | 1.212 | 3.386 | 2.651 | 2.017 | 1.693 | 1.184 |
| 350 | 3.612 | 2.827 | 2.151 | 1.805 | 1.263 | 3.515 | 2.752 | 2.094 | 1.757 | 1.229 | 3.465 | 2.713 | 2.064 | 1.732 | 1.211 | 3.384 | 2.649 | 2.016 | 1.692 | 1.183 |
| 360 | 3.607 | 2.823 | 2.148 | 1.803 | 1.261 | 3.512 | 2.749 | 2.092 | 1.755 | 1.228 | 3.463 | 2.710 | 2.062 | 1.731 | 1.210 | 3.383 | 2.648 | 2.015 | 1.691 | 1.183 |
| 370 | 3.602 | 2.820 | 2.146 | 1.801 | 1.259 | 3.508 | 2.747 | 2.090 | 1.754 | 1.227 | 3.460 | 2.708 | 2.061 | 1.730 | 1.210 | 3.382 | 2.647 | 2.014 | 1.690 | 1.182 |
| 380 | 3.598 | 2.816 | 2.143 | 1.798 | 1.258 | 3.505 | 2.744 | 2.088 | 1.752 | 1.225 | 3.458 | 2.707 | 2.059 | 1.728 | 1.209 | 3.380 | 2.646 | 2.013 | 1.690 | 1.182 |
| 390 | 3.593 | 2.813 | 2.140 | 1.796 | 1.256 | 3.502 | 2.742 | 2.086 | 1.751 | 1.224 | 3.455 | 2.705 | 2.058 | 1.727 | 1.208 | 3.379 | 2.645 | 2.013 | 1.689 | 1.181 |
| 400 | 3.589 | 2.809 | 2.138 | 1.794 | 1.255 | 3.499 | 2.739 | 2.084 | 1.749 | 1.223 | 3.453 | 2.703 | 2.057 | 1.726 | 1.207 | 3.378 | 2.644 | 2.012 | 1.688 | 1.181 |
| 425 | 3.579 | 2.802 | 2.132 | 1.789 | 1.251 | 3.492 | 2.734 | 2.080 | 1.746 | 1.221 | 3.448 | 2.699 | 2.054 | 1.723 | 1.205 | 3.375 | 2.642 | 2.010 | 1.687 | 1.180 |
| 450 | 3.570 | 2.795 | 2.127 | 1.785 | 1.248 | 3.486 | 2.729 | 2.077 | 1.743 | 1.219 | 3.443 | 2.695 | 2.051 | 1.721 | 1.204 | 3.372 | 2.640 | 2.009 | 1.686 | 1.179 |
| 475 | 3.562 | 2.788 | 2.122 | 1.781 | 1.245 | 3.481 | 2.725 | 2.073 | 1.740 | 1.217 | 3.438 | 2.692 | 2.048 | 1.719 | 1.202 | 3.370 | 2.638 | 2.007 | 1.685 | 1.178 |
| 500 | 3.555 | 2.783 | 2.117 | 1.777 | 1.243 | 3.475 | 2.721 | 2.070 | 1.737 | 1.215 | 3.434 | 2.689 | 2.046 | 1.717 | 1.201 | 3.368 | 2.636 | 2.006 | 1.683 | 1.177 |
| 525 | 3.548 | 2.777 | 2.113 | 1.773 | 1.240 | 3.471 | 2.717 | 2.067 | 1.735 | 1.213 | 3.431 | 2.686 | 2.043 | 1.715 | 1.199 | 3.366 | 2.635 | 2.005 | 1.682 | 1.177 |
| 550 | 3.541 | 2.772 | 2.109 | 1.770 | 1.238 | 3.466 | 2.713 | 2.065 | 1.733 | 1.212 | 3.427 | 2.683 | 2.041 | 1.713 | 1.198 | 3.364 | 2.633 | 2.004 | 1.681 | 1.176 |
| 575 | 3.535 | 2.767 | 2.106 | 1.767 | 1.236 | 3.462 | 2.710 | 2.062 | 1.731 | 1.210 | 3.424 | 2.680 | 2.039 | 1.712 | 1.197 | 3.362 | 2.632 | 2.003 | 1.681 | 1.175 |
| 600 | 3.530 | 2.763 | 2.102 | 1.764 | 1.234 | 3.458 | 2.707 | 2.060 | 1.729 | 1.209 | 3.421 | 2.678 | 2.038 | 1.710 | 1.196 | 3.360 | 2.631 | 2.002 | 1.680 | 1.175 |
| 625 | 3.525 | 2.759 | 2.099 | 1.762 | 1.232 | 3.455 | 2.704 | 2.058 | 1.727 | 1.208 | 3.418 | 2.676 | 2.036 | 1.709 | 1.195 | 3.359 | 2.629 | 2.001 | 1.679 | 1.174 |
| 650 | 3.520 | 2.755 | 2.096 | 1.759 | 1.230 | 3.451 | 2.702 | 2.056 | 1.725 | 1.207 | 3.416 | 2.674 | 2.034 | 1.707 | 1.194 | 3.357 | 2.628 | 2.000 | 1.678 | 1.174 |
| 675 | 3.515 | 2.752 | 2.094 | 1.757 | 1.229 | 3.448 | 2.699 | 2.054 | 1.724 | 1.205 | 3.413 | 2.672 | 2.033 | 1.706 | 1.193 | 3.356 | 2.627 | 1.999 | 1.678 | 1.173 |
| 700 | 3.511 | 2.748 | 2.091 | 1.755 | 1.227 | 3.445 | 2.697 | 2.052 | 1.722 | 1.204 | 3.411 | 2.670 | 2.032 | 1.705 | 1.192 | 3.355 | 2.626 | 1.998 | 1.677 | 1.173 |
| 725 | 3.507 | 2.745 | 2.089 | 1.753 | 1.226 | 3.442 | 2.694 | 2.050 | 1.721 | 1.203 | 3.408 | 2.668 | 2.030 | 1.704 | 1.192 | 3.354 | 2.625 | 1.998 | 1.676 | 1.172 |
| 750 | 3.503 | 2.742 | 2.086 | 1.751 | 1.225 | 3.439 | 2.692 | 2.049 | 1.719 | 1.202 | 3.406 | 2.667 | 2.029 | 1.703 | 1.191 | 3.352 | 2.624 | 1.997 | 1.676 | 1.172 |
| 800 | 3.495 | 2.736 | 2.082 | 1.747 | 1.222 | 3.434 | 2.688 | 2.046 | 1.717 | 1.201 | 3.402 | 2.663 | 2.027 | 1.701 | 1.189 | 3.350 | 2.623 | 1.996 | 1.675 | 1.171 |
| 850 | 3.489 | 2.731 | 2.078 | 1.744 | 1.220 | 3.430 | 2.685 | 2.043 | 1.714 | 1.199 | 3.399 | 2.661 | 2.025 | 1.699 | 1.188 | 3.348 | 2.621 | 1.994 | 1.674 | 1.171 |
| 900 | 3.483 | 2.726 | 2.075 | 1.741 | 1.218 | 3.426 | 2.682 | 2.040 | 1.712 | 1.198 | 3.396 | 2.658 | 2.023 | 1.697 | 1.187 | 3.347 | 2.620 | 1.993 | 1.673 | 1.170 |
| 950 | 3.477 | 2.722 | 2.071 | 1.738 | 1.216 | 3.422 | 2.679 | 2.038 | 1.710 | 1.196 | 3.393 | 2.656 | 2.021 | 1.696 | 1.186 | 3.345 | 2.619 | 1.992 | 1.672 | 1.169 |
| 1,000 | 3.472 | 2.718 | 2.068 | 1.736 | 1.214 | 3.418 | 2.676 | 2.036 | 1.709 | 1.195 | 3.390 | 2.654 | 2.019 | 1.695 | 1.185 | 3.344 | 2.617 | 1.992 | 1.671 | 1.169 |
| ∞ | 3.291 | 2.576 | 1.960 | 1.645 | 1.150 | 3.291 | 2.576 | 1.960 | 1.645 | 1.150 | 3.291 | 2.576 | 1.960 | 1.645 | 1.150 | 3.291 | 2.576 | 1.960 | 1.645 | 1.150 |

TABLE 2.2

COEFFICIENTS TO FACILITATE THE USE OF FORMULA (6) FOR COMPUTING $K$ WHEN $N$ IS
LARGE

| Coefficient | Value | $\gamma = 0.75$ | $\gamma = 0.90$ | $\gamma = 0.95$ | $\gamma = 0.99$ |
|---|---|---|---|---|---|
| Term in $\dfrac{1}{\sqrt{N}}$ .......... | $\dfrac{x_\gamma}{\sqrt{2}}$ | 0.4769 | 0.9062 | 1.1631 | 1.6450 |
| Term in $\dfrac{1}{N}$ ........... | $\dfrac{5x_\gamma{}^2 + 10}{12}$ | 1.0229 | 1.5177 | 1.9606 | 3.0883 |

## 3. CONSTRUCTION OF THE TABLE

### 3.1. Formula

The formula for $K$ used in Table 2.1 is*

$$K = r \sqrt{\frac{n}{\chi_\gamma^2}} \tag{13}$$

where $N$ is the number of observations, $P$ is the proportion of the distribution to be included within the tolerance limits, $\gamma$ is the confidence coefficient, $r$ is a function of $1/\sqrt{N}$ and $P$ defined by

$$\frac{1}{\sqrt{2\pi}} \int_{\frac{1}{\sqrt{N}}-r}^{\frac{1}{\sqrt{N}}+r} e^{-\frac{t^2}{2}} \, dt = P \tag{14}$$

and $\chi_\gamma^2$ is defined by

$$P(\chi^2 > \chi_\gamma^2) = \gamma \tag{15}$$

in terms of the chi-square distribution for $n$ degrees of freedom, where

$$n = N - 1. \tag{16}$$

### 3.2. Basic Network of Points

$K$ was computed directly from formula (13) for all values of $n$ (that is, of $N - 1$) from 1 to 30, for all even values of $n$ from 30 to 60, for values of $n$ from 60 to 200 at intervals of 10, and for values of $n$ from 199 to 999 at intervals of 50. For other values of $n$, values of $K$ were interpolated into this basic network of points by means of 4-point Lagrangian interpolatory coefficients taken from the table published by the New York Mathematical Tables Project.

### 3.3. Percentage Points of $\chi^2$

For all values of $n$ from 1 to 30 and for values from 30 to 100 at intervals of 10, the percentage points of $\chi^2$ were obtained from Thomp-

* This formula was developed in the Statistical Research Group by A. Wald and J. Wolfowitz; see their paper cited in Sec. 4.

son's table.   Since for even $n$ the probability of exceeding any given value $\chi^2$ is given by the partial sum of a Poisson series, the required values between $n = 30$ and $n = 60$ (except $n = 40$ and $n = 50$) were obtained by inverse interpolation in Molina's table* using Salzer's inverse interpolation coefficients.   The percentage points of $\chi^2$ for $n > 100$ were found from a Cornish-Fisher type of expansion applied to the $\chi^2$ distribution.†

### 3.4. Computation of $r$

To compute $r$, Newton's method was applied to the function

$$G(r) = \frac{1}{\sqrt{2\pi}} \int_{\frac{1}{\sqrt{N}}-r}^{\frac{1}{\sqrt{N}}+r} e^{-\frac{t^2}{2}} dt - P = 0 \tag{17}$$

The NYMTP tables of the probability integral are conveniently arranged for this, since the ordinates of the normal curve are tabulated beside the integral, and forward interpolation in both the integral and the ordinate may be expressed simply in terms of the ordinate.

### 4. REFERENCES

BOWKER, ALBERT H.: "Computation of Factors for Tolerance Limits on a Normal Distribution When the Sample Is Large," *Annals of Mathematical Statistics*, Vol. 17 (1946), pp. 238–240.

GOLDBERG, HENRY, and LEVINE, HARRIET: "Approximate Formulas for the Percentage Points and Normalization of $t$ and $\chi^2$," *Annals of Mathematical Statistics*, Vol. 17 (1946), pp. 216–225.

Mathematical Tables Project, Work Projects Administration for the City of New York, Federal Works Agency: *Tables of Probability Functions*, Vol. 2, National Bureau of Standards, Washington, D. C., 1942, 344 + xxi pp.

————: *Tables of Lagrangian Interpolation Coefficients*, Columbia University Press, New York, 1944, 392 + xxxvi pp.

MOLINA, E. C.: *Poisson's Exponential Binomial Limit*, D. Van Nostrand Company, Inc., New York, 1942, 47 pp.

PAULSON, EDWARD: "A Note on Tolerance Limits," *Annals of Mathematical Statistics*, Vol. 14 (1943), pp. 90–93.

ROBBINS, HERBERT: "On Distribution-free Tolerance Limits in Random Sampling," *Annals of Mathematical Statistics*, Vol. 15 (1944), pp. 214–216.

SALZER, H. E.: "Table of Coefficients for Inverse Interpolation with Advancing Differences," *Journal of Mathematics and Physics*, Vol. 23 (1944), pp. 75–102.

SCHEFFÉ, H., and TUKEY, J. W.: "A Formula for Sample Sizes for Population Tolerance Limits," *Annals of Mathematical Statistics*, Vol. 15 (1944), p. 217.

* In Molina's notation

$$P\left(\frac{n}{2}, \frac{\chi_\gamma^2}{2}\right) = 1 - \gamma$$

† The application of the Cornish-Fisher method to the $\chi^2$ distribution was carried out in the Statistical Research Group by Henry Goldberg and Harriet Levine; see their paper cited in Sec. 4.

SHEWHART, WALTER A., with the editorial assistance of DEMING, W. EDWARDS: *Statistical Method from the Viewpoint of Quality Control*, Graduate School, U.S. Department of Agriculture, Washingon, D. C., 1939, 155 + ix pp.

THOMPSON, CATHERINE M.: "Table of Percentage Points of the $\chi^2$ Distribution," *Biometrika*, Vol. 32 (1941), pp. 188–189.

WALD, ABRAHAM: "Setting of Tolerance Limits When the Sample Is Large," *Annals of Mathematical Statistics*, Vol. 13 (1942), pp. 389–399.

——: "An Extension of Wilks' Method of Setting Tolerance Limits," *Annals of Mathematical Statistics*, Vol. 14 (1943), pp. 45–55.

——, and WOLFOWITZ, J.: "Tolerance Limits for a Normal Distribution," *Annals of Mathematical Statistics*, Vol. 17 (1946), pp. 208–215.

WILKS, S. S.: "Determination of Sample Sizes for Setting Tolerance Limits," *Annals of Mathematical Statistics*, Vol. 12 (1941), pp. 91–96.

——: "Statistical Prediction with Special Reference to the Problem of Tolerance Limits," *Annals of Mathematical Statistics*, Vol. 13 (1942), pp. 400–409.

# CHAPTER 3

## MULTIVARIATE QUALITY CONTROL, ILLUSTRATED BY THE AIR TESTING OF SAMPLE BOMBSIGHTS

.

.

.

### This chapter was prepared by
### Harold Hotelling, University of North Carolina
.

# CHAPTER 3

# MULTIVARIATE QUALITY CONTROL, ILLUSTRATED BY THE AIR TESTING OF SAMPLE BOMBSIGHTS

## 1. INTRODUCTION

### 1.1. Type of Problem

Systematic plans of sampling for acceptance inspection or quality control are for the most part based on simple features of the individual articles inspected. Usually each article is classified merely as defective or nondefective. In a minority of cases there is a finer classification involving measurement of continuous variates. In the inspection of simple articles, including component parts of complicated assemblies, inspection by attributes is considered appropriate; but in the inspection of complicated assemblies, especially when inspection is costly, measurements are apt to be advantageous.* Furthermore, the performance of complicated assemblies is usually a function of several different variables. When several quality characteristics are measured, there arises a problem of combining the several measurements on each article into a single measure ("score" or "index") of the over-all performance of the article.†

In the absence of a formally defined procedure, methods of a more or less subjective character for combining the different measurements to form a judgment of degree of acceptability are bound to develop. The simplest such method is to apply customary practices to each variate separately and to reject whenever the usual rules applied to any one variate alone lead to a verdict of rejection. This procedure, however, has the effect of changing drastically the probabilities, both of accepting what is bad and of rejecting what is good. For example, if a sampling scheme is designed to give a probability of 0.05 of rejecting a lot of a certain quality on the basis of measurements of one variate, and if this scheme is applied to each of two variates measured on the same articles, then if the two variates are independently distributed, rejection on the basis of one or the other variate will take place in 9.75 per cent of submitted lots of the given quality instead of 5 per cent as originally contemplated. The discrepancy between the actual and the intended probabilities of rejection will increase rapidly with the number of independent variates when such a rule is followed.

* For a further discussion of this point, see Chap. 1, Sec. 1.
† For a further discussion of this point, see Chap. 1, Sec. 10.1.

An alternative rule, erring in the opposite direction, would be to accept any lot for which the univariate scheme applied to *any one* of the variates calls for acceptance. Thus rejection would require that *all* the variates take such values as to call for rejection. For two independent variates, a probability of rejection intended to be 0.90 would actually in such a case be only 0.81.

These probabilities can be adjusted so as to become equal to 0.05, or such level as is chosen, by altering the acceptance level for each variate separately. However, this brings in additional difficulties. The variates may not be mutually independent, and then calculations such as the foregoing must be altered to take into account the multivariate distribution. Furthermore, it will often not be known whether they are independent or not; or if they are mutually dependent, the character of the dependence may be known only imperfectly. Thus the correlation coefficient may have to be estimated from a preliminary sample of size so small as to leave its exact value somewhat uncertain. Any acceptance probabilities based on such a correlation coefficient will then likewise be uncertain.

Rules such as those suggested above have another defect. An article close to the margin of acceptability with respect to one variate may well be marked for acceptance or rejection on the basis of the other variates involved. Unusual excellence in one respect may often atone for a slight departure in another way from what would otherwise be considered satisfactory. A formula is then needed for combining the measures in the several dimensions into a single measure of excellence, or of adherence to specifications or to a desired ideally uniform set of dimensions. Such a formula should give a measure of departure from the ideal that will increase with a departure in any dimension. It should lend itself to an exact interpretation in terms of probability, taking account of the necessarily limited size of any preliminary sample used to determine the parameters entering into the formula.

Formulas of this type are here developed on the basis of an expression known as the generalized Student ratio $T$, introduced in 1931 for the testing of hypotheses about the location of the means of multivariate distributions and closely related to the exponent in the multivariate normal distribution.* If errors in two or more dimensions were mutually uncorrelated and of equal variance in the sample representing past experience, the value of $T$ for a new individual would be directly proportional to the distance from the point whose rectangular coordinates represent the ideal to the point representing the actual measurements. Thus, for a bomb dropped on a target, $T$ would in such a case be simply

---

* See HOTELLING, "The Generalization of Student's Ratio." (For full citations, see Sec. 6.)

the distance of the bomb from the center of the target, expressed in terms of a unit equal to the standard deviation found in the earlier experience. Actually, the variances of range errors in bombing are appreciably less than those of deflection errors.*   Consequently, the expression used for $T^2$ has different coefficients for the two squared terms.   In addition, it has a term in the product of range and deflection errors, the small coefficient of which results from a slight correlation between the two kinds of error in the available experience.   This $T^2$ serves as a measure of the deviation of the particular bomb from the center of the target, a measure more accurately interpretable in terms of probability than is the actual distance.   By adding the values of $T^2$ for all the bombs dropped with a particular bombsight, we have a measure useful in judging the accuracy of that bombsight.   There are also other ways of combining these $T$'s to throw light on particular features.

A great advantage of all these $T$ statistics is that, in samples from normal distributions, they have probability distributions that can be determined *exactly*.   Many statistics have distributions involving what may be called "nuisance parameters," that is, parameters which are not known exactly from general considerations or basic scientific constants but have to be estimated, with more or less uncertainty, from fallible observations.   This is true, for example, of the actual distance of a bomb from a target, whose distribution depends not only on the mean value of such distances but also, and in a complicated way, on the *direction* of the deviation.   The parameters specifying the distribution of the direction—which is really irrelevant—are nuisance parameters in the distribution of the actual distance, confusing its interpretation as a measure of accuracy of the bomb.   But the $T$ statistics have been constructed in such a way that there are no nuisance parameters involved in their distributions, at least as long as we are dealing with data from a fundamental distribution of the normal form.

It is to be expected that these $T$ statistics will have many applications in sampling inspection and in other fields.   In particular, wherever a manufactured article is to be judged, not on the basis of measurement of parts, but on the basis of final performance as measured by several variables, there seems to be room for the utilization of such methods.   A case of this kind, which is here considered in detail and for which the methods presented here were originated, is that of air testing sample bombsights. It is presented here as a precedent that will be useful in other situations in which final performance is determined by multiple measurements.

* An error in range is an error in the direction of the airplane's heading at the time of release.   An error in deflection is an error in a direction perpendicular to the plane's heading.

The mathematical theory is not given in full in this chapter, but the methods recommended will, it is believed, be fully intelligible on the basis of the material presented here.*

Besides the $T$ distribution, several other innovations are here suggested, both graphic devices and methods of applying to statistical quality control some of the ingenious devices and concepts that R. A. Fisher has introduced into the design of experiments.

### 1.2. Problems of Air Testing Bombsights

The dropping of bombs from airplanes for the purpose of testing bombsights submitted by manufacturers is a phase of quality control. The art and the literature of quality control, so greatly expanded of late, are to a certain extent applicable to this bombing procedure.   Yet test bombing presents several novel features that make the situation very different from those heretofore dealt with in the literature and usual practice of quality control.   These peculiar features, which make this no ordinary problem and render the usual methods inadequate and to some extent inappropriate, are the following:

(1) In contrast to the simple articles to which routine mass sampling inspection is usually applied, with at most one or two numbers recorded for each article tested, and often merely a "good" or "bad" verdict, the bombsight is a highly complex instrument with a multiplicity of parts, each of which is subject to various possible defects.

(2) The process of putting together these parts in selective assembly, matching an oversize part with an undersize part to fit in with it, grinding and adjusting when necessary, is a series of jobs requiring great skill, whose results are to be judged by the performance of the finished instrument rather than by measurement of parts.   Performance, however, is a complicated thing.   If quality is to be measured by performance and expressed by one or a few numbers, caution must be exercised in the choice of these numbers and their interpretation.

(3) Air testing is only one of a series of tests and inspections to which a bombsight and its parts are subjected.   It is the final step, and an exceptionally costly one.   For elimination of defects the main reliance is properly placed on earlier and cheaper steps.   A great deal of careful screening out of defectives has taken place prior to the final stage of air testing, with which alone this chapter is concerned.

(4) The errors of test bombing are not by any means due entirely to defects in the sight.   In spite of all attempts at standardization there are bound to be variations in the weather, the altitude from which the bombs are dropped, the bombs themselves, the airplane, the auxiliary equip-

* It is planned to present the mathematical theory in the *Annals of Mathematical Statistics.*

ment, and the personnel engaged in flying, bombing, and observing the results. The variance in the results of the bombing that results from causes having nothing to do with the quality of the bombsight indicates that, while valuable statistical inferences are possible from large numbers of cases, it would be a dubious procedure to take any drastic action on the basis of the four or six bombs usually dropped with a sight, except perhaps in the most extreme cases.

(5) When bombing with a particular sight shows unsatisfactory results, the action ordinarily taken is not the rejection of the sight, but its adjustment or repair.

(6) Whereas the usual sampling inspection for quality control is carried out very cheaply for each single inspection by a workman of moderate skill using some simple instrument, the dropping of a test bomb requires a pilot, a bombardier, an airplane with gasoline and maintenance, an airfield, a target area, stations for observing the point of impact and personnel to man them, and finally, plotting and computations.

The function of air testing is clearly not the elimination of individual bad sights or of bad lots of 10 from which the sights tested are taken. This is done far more efficiently by inspection of parts and by shop testing of the finished bombsight, both by the manufacturer and by the government representatives at the factory, as well as by the shop testing of the sample sent to the government proving ground. The principal functions which air testing is capable of fulfilling appear rather to be the following:

(1) Keeping a running record of quality designed to compare the work of different productive units, and of the same unit at different times.

(2) Detection of irregularities, trends in quality, and deviations from customary levels in the average performance of sights produced over a period.

(3) Exploration of possible types of defects, either in design or in production, not previously recognized.

(4) By experiments carefully designed so as not to conflict with the other purposes of air testing, to obtain judgments on differing models of sights, on auxiliary equipment, on bombing methods, and on personnel in a manner more efficient than separate experiments carried out solely to obtain these judgments.

The high cost and uncertainty of air testing, with the relative accuracy and low cost of shop testing, indicate that only a very small number of bombsights should be air-tested. However, it does appear that some air testing should go on, for reasons outside those usual in sampling inspection and to an extent that cannot be determined by calculation according to customary methods. In the specific, though flexible, plan suggested in Sec. 2, the amount of routine bombing called for is one-twentieth of that under procedures in effect in 1943. The exact proportion suitable

cannot be derived by calculation from the proportion of defects, which could not be ascertained reliably from the record because of the virtual absence of defects serious enough to cause rejection.   The fraction suggested represents what is in the author's view a reasonable proportion of bombsights to be tested if the improved statistical methods described herein are utilized.   The plan can easily be modified, in accordance with Sec. 3, to fit any other proportion considered appropriate.

In much quality-control work where individual inspection is cheap, inefficient statistical formulas are used frankly and deliberately when they are easy to calculate and to comprehend.   The saving in computation represented by using the mean deviation instead of the standard deviation with many small batches may in some circumstances offset the increase of approximately one-eighth in the amount of inspection necessary to get the same accuracy.   The range of a sample (that is, the difference between the greatest and least of the values) is often taken as a measure of dispersion when it is felt that the inability of factory employes to understand any more sophisticated measure warrants the cost of the additional observations required to obtain the necessary accuracy.

Considerations of this kind should have no place where the unit cost is as great as that of dropping a bomb.   Information acquired at such great expense should not be wasted by the use of inefficient statistical methods of combining the observations.   One object of this chapter is to set forth the most appropriate methods of combining the observations of the bomb coordinates and of making inferences from them.   Some modifications of the usual kinds of control charts are introduced.   Furthermore, the contribution of modern statistical methods to the improvement of accuracy is not limited to formulas for combining the observations after they are made.   Certain precautions, such as proper randomization and replication, can be arranged only *before* the observations are made.   The application of these principles to air-testing sample bombsights by a plan in accordance with modern principles of experimental design is another leading purpose of this chapter.

Section 2 consists of a specific plan for the selection of sample bombsights, their air testing, the calculations to be made from the results, a series of control charts of new types, and some simple considerations relevant to interpretation of the results and to action.

Since Sec. 2 is easily adaptable to changed conditions and modifications of the plan, formulas are given in Sec. 3 for calculating new control limits and other needed quantities under conditions more general than those of Sec. 2.   Provision is here made for changing the number of bombs dropped per flight, the number of flights with each sight, the number of sights to be taken from a lot for air testing, and the base period with which comparison is to be made.

While Secs. 2 and 3 consist principally of simple rules for operating and for amending the sampling plan, with little mathematics or argumentation, Sec. 4 gives the background and reasons for the plan. The chief mathematical discussion has, however, been confined to a technical paper, where the mathematical reader will find a more precise account of some of the new ideas presented here for the first time. These ideas are applicable to a wide range of problems, not only in quality control but also in connection with bombing and firing in general, with mental testing, and with other problems in which two or more normally distributed variates are involved.

Section 5 is an analysis of the results of dropping 260 ore-filled bombs. Measures of dispersion obtained from these bombs are substituted into the formulas of Sec. 3 to obtain the control limits specified in the plan of Sec. 2, together with coefficients for combining range and deflection errors and other needed quantities. For any sampling plan, a base period is needed to get the plan started, and this is provided by the material analyzed in Sec. 5. This section also contains as by-products a specimen rating of equipment and personnel involved in that bombing, essentially to show how such ratings can be made, and a new simplified technique of analysis of variance with unequal class frequencies developed for the purpose of making these ratings.

This chapter includes control charts having several novel features, which it is believed will also be found useful in other control-chart work. One is the choice of the functions combining different measures of accuracy, here range and deflection errors, to be plotted. These functions are invariant under transformations of coordinates and hence have probability distributions capable of exact determination. Another feature of the control charts here presented is the provision of *yardsticks*, for visual comparison of different plotted values, with a definite probability interpretation.

In fixing the control limits corresponding to specified probabilities, account is here taken of the limited number and consequent sampling fluctuations of the observations in the base period used for comparison. It is to be feared that in many control charts heretofore in use there has been an unconscious distortion of the probabilities because of neglect of this consideration, with a bias uniformly in the direction of exaggerating the significance of moderately large deviations from norms. Use of formulas such as those of Sec. 3 in industrial quality control should correct this error.

## 2. PROPOSED SAMPLING PLAN

### 2.1. Selection of Sample Sights and Arrangement of Bombing

(1) Whenever a new factory goes into production, each of the first 40 bombsights produced and passing the usual shop tests of the govern-

ment inspectors, both at the factory and at the proving ground, is to be tested.   If all these are found satisfactory, the remainder of the plan is to be applied to subsequent production, subject to withdrawal with reversion to the status of a new factory if later a series of unacceptable sights is offered.   (Since the last contingency has not hitherto occurred, accurate formulation of a definite rule for such reversion must await the accumulation of this undesirable kind of experience.)

(2) At each factory that has been approved after the first 40 bombsights, the sights of each model are to be numbered consecutively in the order of manufacture in each of two series.   One series, designated $A$, shall consist of those sights approved without unusual delay by the government inspector at the factory on first delivery by the manufacturer. The other series, designated $B$, shall consist of those sights first rejected by the government inspector and later approved after defects have been remedied, together with any other approved sights whose approval has for any reason been delayed.   To each sight are to be affixed both the usual serial number representing original order of manufacture and the letter $A$ or $B$, with the number within the $A$ or $B$ series.   The subsequent procedures are to be applied separately to the $A$ and $B$ series for each factory and for each model.

(3) Each series is to be divided into successive lots of 20, strictly in the order of original manufacture.

(4) From each lot of 20, two sights are to be selected by an objective chance procedure and sent to the proving ground.   No element of personal judgment or of rating on inspection should enter into this selection, which may be made in the following manner: the procurement authorities will send to the chief inspector at the factory a series of opaque sealed envelopes marked on the outside with the first and last serial numbers of the successive batches.   An envelope is to be opened only after the inspection at the factory of all the sights in a batch has been completed and report made.   Within the envelope will be found two numbers between 1 and 20, inclusive, prepared on a strictly chance basis, with equal probabilities for all twenty numbers.   This must be done through actual utilization of a chance mechanism or of a table of random numbers, * without human discretion other than in the selection of the mechanism.   One obvious procedure would be to mark twenty playing cards with the numbers 1 to 20, and after thoroughly shuffling to draw out two of them at random.

(5) The remaining 18 sights of each lot are not to be held at the factory until completion of the air testing of the sample but should be

---

· * For example, Fisher and Yates, *Statistical Tables*, Table XXXIII; Kendall and Smith, *Random Sampling Numbers;* Tippett, *Random Sampling Numbers.*

sent directly to service, subject to later recall or repair in the unlikely event that the sample is found seriously defective.

(6) Upon arrival at the proving ground, the sample sights are all to be shop-tested and adjusted as at present, with such further types of shop test as may be developed.

(7) By an objective random procedure like that described in Paragraph (4) above, lots are to be selected for the air testing of the two sights representing the lot at the proving ground. It is suggested that one-fortieth of the lots produced by each manufacturer be selected by chance. (It is likely that this proportion should be reduced as the improved methods of shop testing prove their value.)

(8) Each of the two sights representing each of the lots selected in accordance with Paragraph (7) is to be taken up in an airplane twice to be used for dropping bombs.

(9) The routine flight testing is to be arranged in periods, each consisting of four consecutive days, or four days as nearly consecutive as possible even though there may be breaks on account of Sundays or special circumstances. The number of test flights to be made per day should first be determined and will here be denoted by $k$. This same number of flights $k$ should be made on each day of the 4-day period, though it is not essential that $k$ should be the same for different periods.

(10) The sights to be tested in a 4-day period are to be the two sample sights from each of $k$ lots. Each sight is to be tested on two different days of the period, and each lot is to be represented on all four days. Subject to these conditions, the assignment of the particular sight of a lot to be tested on a specific day is to be decided by objective chance, for example, by flipping a coin. The $k$ lots represented may be from the same or different factories. They may be of different models and may be of $A$ or $B$ series or both.

(11) Exactly four bombs are to be dropped on each flight. Range and deflection errors are to be measured with respect to the heading of the airplane at the time the particular bomb was dropped.

(12) The numerical values for control limits and yardsticks given in this section are to be used only if the directions in the preceding eleven paragraphs are carried out exactly. If it is decided to revise the plan by dropping a different number of bombs on each flight, or by measuring range and deflection with respect to the true course instead of the heading, the numerical values will need to be amended in accordance with the formulas of Sec. 3. The number of bombs dropped per flight will be denoted by $N$, with the understanding in this section that $N = 4$. No change should be made except as part of an explicit revision of the plan.

(13) All measures of quality based on the bombing results should be algebraic functions of the coordinates of the bombs with respect to the

target, taken with their proper algebraic signs. Absolute values and mean deviations are not algebraic functions and should not be employed in this connection. Use of criteria based on mean deviations requires more bombing to procure comparable accuracy than the formulas used herein. Furthermore, in contrast to the $T$ measures here recommended, no combination of mean deviations has an exactly known distribution.

## 2.2. Control Charts

**2.2.1. General Plan.**—Seven control charts, illustrated by Figs. 3.1 to 3.7,* are to be kept for each factory and model and for the $A$ and $B$ series. On these charts certain indicators of quality of successive lots are plotted, strictly in the order of production. Entries on these charts should be made daily as soon as the data become available. The charts should be posted conspicuously at the proving ground where they will be seen by all concerned with the testing. Copies of them should be sent weekly to the procurement authorities, to the particular manufacturer concerned, and to the government inspector at the factory. The charts should all be drawn with the same horizontal scale, representing the order of the lot represented, and should be placed one above another with a clear indication of the boundaries between them when this is practicable. Space should be left at the bottom for comments on exceptional cases and a notation of action taken with respect to such cases.

The quantities represented on the charts, from top to bottom in order, will be designated by $T_O$, $T_D$, $T_M$, $s_r$, $\bar{r}$, $s_d$, $\bar{d}$. The highest of these, $T_O$, is the most comprehensive measure of the bombing errors associated with the sights under consideration, and anyone who has time to consider only one measure should watch $T_O$. The next two break down the overall error into a part representing deviations from the mean point of impact (MPI) of the bombs on a flight and a part representing deviations of the MPI from the target. All these first three measures are combinations of range errors with deflection errors. Of the remaining four quantities charted, two refer to range errors alone and two to deflection errors alone, two to the MPI and two to deviations from the MPI.

Any general tendency to slipshodness and poor work in a factory is

---

* These illustrative charts were made up from the actual bombing data discussed in Sec. 5 and are thus mutually consistent. Though these records of flights were all derived from different sights, they were arbitrarily grouped into consecutive pairs, representing fictitious single sights, and these pairs were again grouped into consecutive pairs representing samples from hypothetical lots, just as if the plan here presented had been in effect. The last flight was omitted so as to leave 64 flights, a multiple of 4, as is required by this plan. The result is to give a reasonably good illustration of the kind of charts that can be expected in test bombing of the kind carried out with bombsights of the same general quality as the ones then tested.

more likely to be observed in rising curves in the upper charts, approaching and passing the control limits, whereas light on specific faults is to be sought by casting the eye downward gradually to the lower charts.

The seven, quantities to be charted will be defined in the following paragraphs, beginning with the simplest. They will therefore be in a different order from that above. The actual ordinates are not proportional to the quantities mentioned above, except $\bar{r}$ and $\bar{d}$, but to their logarithms. Consequently, arithlog (that is, semilogarithmic) paper should be used for the top five charts to save the trouble of looking up logarithms. Also, the vertical scales on these five top charts are to be marked on the left with numerical values equal to the squares of the $T$'s and $s$'s since, as will be seen later, this avoids the labor of extracting square roots. The $T$'s and $s$'s themselves are given on vertical scales at the right of these charts.

We shall use $r$ and $d$ as symbols for range and deflection errors, respectively, in mils,* and indicate by $S$ the sum over the $N$ ($=4$) bombs in a flight of the values of the expression following $S$.

All the numerical coefficients, yardstick lengths, and control limits herein specified should be revised periodically as further experience accumulates; they will especially need to be revised after any change in technique has been in effect long enough to provide a sufficient basis in experience. Formulas for revision will be found in Sec. 3.

**2.2.2. Control Chart for Mean Range.**—The control chart for $\bar{r}$, illustrated by Fig. 3.1, represents $\bar{r}_F = Sr/N$ for each flight by a cross or a circle. For each lot the values of $\bar{r}_F$ found from the two flights of the sight produced earlier than the other are represented by circles, while those for the sight produced later are represented by crosses. The mean of these four values of $\bar{r}_F$ is $\bar{r}_L$, the mean of the range errors for the lot. The points representing $\bar{r}_L$ for successive lots are connected by straight lines.

Control limits are shown as two pairs of horizontal lines. The probability that a lot of sights constituting a random sample from the same population as the sights tested in the base period, with the same distribution of errors of all kinds as in the test period, should be represented by a point outside the outer control limits is 0.01. The probability that such a lot should yield a value of $\bar{r}_L$ falling outside the inner control limits under the same conditions is 0.05. The upper and lower limits corresponding to each probability are at equal distances, not from the zero line, but from a line representing the range error in the MPI of all the bombs dropped in the base period. This point is discussed in Sec. 3.2, where the formulas and numerical values for the control limits are also

---

* A mil is an angular measure approximately equal to 1/6,400 of 360 degrees or to 1/1,000 of a radian, the exact definition varying from one usage to another.
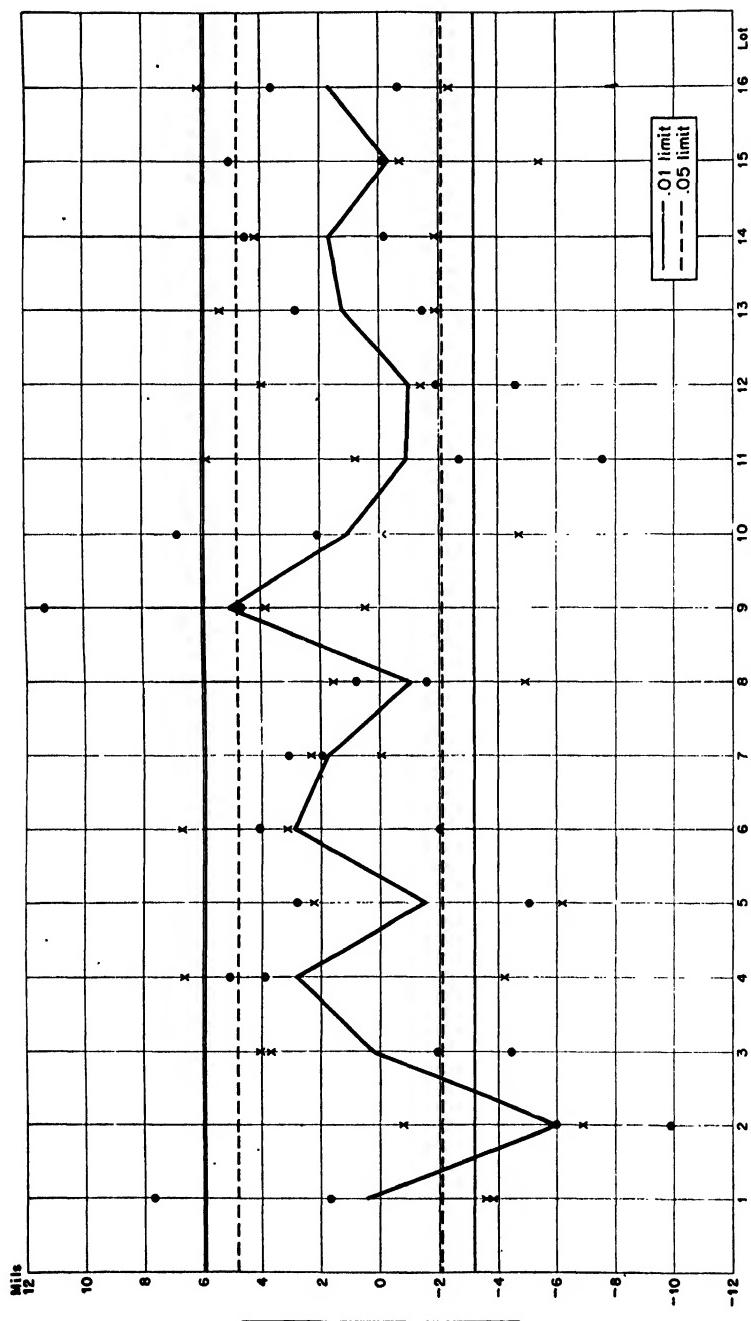
FIG. 3.1.—Control chart for mean range $\bar{r}$.

given.   The 0.01 limits are represented by solid lines at 5.89 and $-3.20$ mils; the 0.05 limits are shown by lines of dashes at 4.80 and $-2.11$ mils.

If the zigzag line should get outside the band marked out by the inner control limits, this fact may be taken as a warning to be alert for any further indication of faults tending to make bombs fall consistently too far short of or too far beyond the target, as the case may be.   If the outer control limits should be surpassed, the warning is more emphatic, since this will happen only once for 100 lots if there are only the same sources of range error as in the base period.   An instance of exceeding the outer control limits should be made the occasion of an investigation at the proving ground to determine the cause.   A definite trend upward or downward in the plotted quantity is another indication of a dangerous condition, even before the inner control limits are reached.   In such contingencies, unless the trend or the excessive value of $\bar{r}_L$ is definitely attributed to faults of the airplane, personnel, or other cause independent of the quality of the sight, the manufacturer should be warned promptly. It is particularly important to give early warning of trends in order that there may be an opportunity to correct gradually developing faults before they become serious.   Action to be taken when control limits are passed is specified further in Sec. 2.3.

The control limits refer to $\bar{r}_L$ but serve a secondary purpose also with respect to $\bar{r}_S$, the mean of the range errors for a sight, represented by a point that is ordinarily not marked but that can be estimated visually from a pair of crosses or of circles.   The control limit having probability 0.01 of being exceeded by $\bar{r}_L$ has a probability approximating 0.20 of being exceeded by $\bar{r}_S$, and that having probability 0.05 for $\bar{r}_L$ has probability practically 0.33 for $\bar{r}_S$, provided the number of bombs dropped in the base period is large, say 200 or more.   Thus the fluctuations of range error in MPI can be followed visually with a fair idea of the relevant probabilities.

To facilitate further the continuing comparison of means of range errors, a yardstick is provided at the left for the difference between the means of two *lots*, a longer yardstick for the difference between the means of two *sights*, and a still longer yardstick for the difference between the means of two *flights*.   The vertical distance between points representing means of any of these types should exceed the length of the corresponding yardstick about once in 20 times if chance alone is operating.   The means for the sights are not shown but are easily inferred as the centers of gravity of the two points representing the flights with the individual sight.   A visual appraisal of the distance between these centers of gravity for the two sights representing a lot, in comparison with the middle yardstick, will usually indicate clearly whether the difference in performance in this respect is unusually great.   An excessive difference should be regarded as a tentative indication of some variation of conditions at the
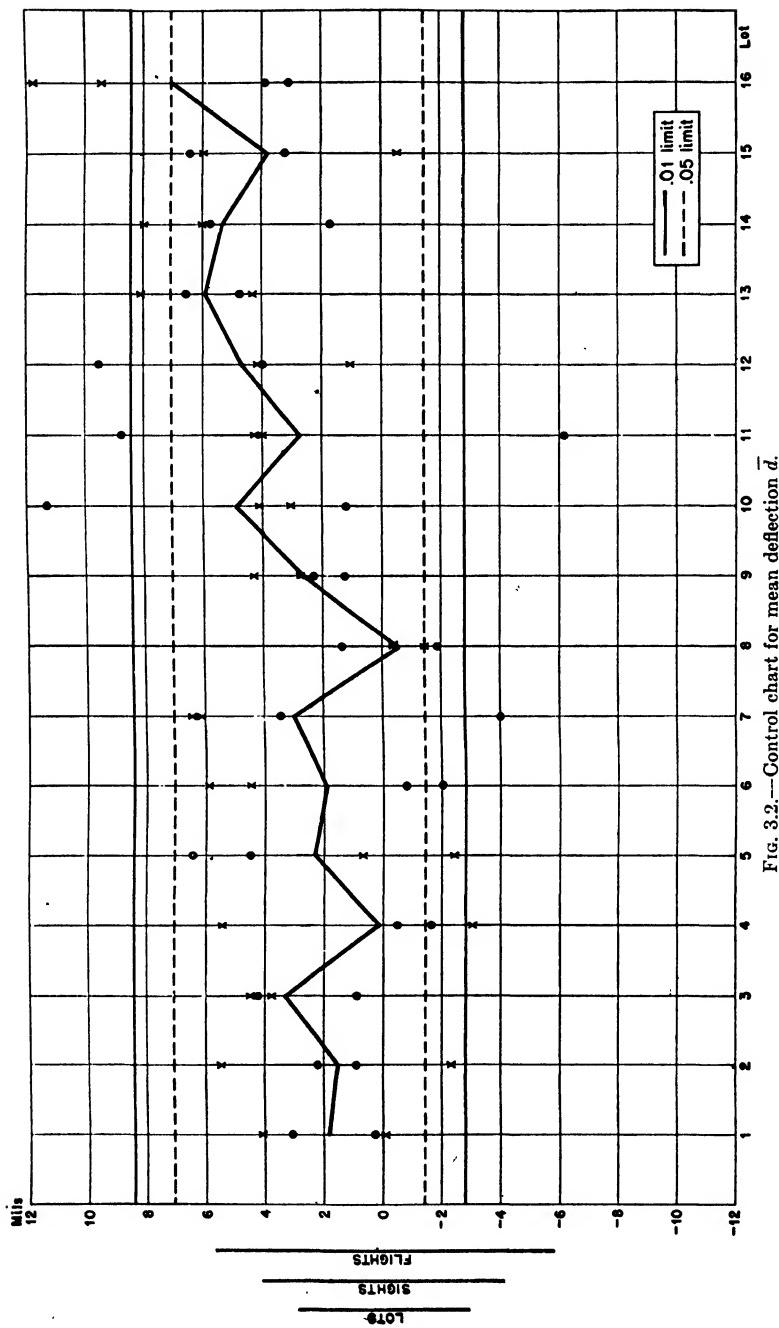
FIG. 3.2.—Control chart for mean deflection $\bar{d}$.

factory during the production of the lot represented by this sample. However, it might also be explained by a difference in weather or personnel conditions between the two days on which one sight was tested and the two days on which the other sight was tested; or by chance alone for every twenty such comparisons if these were to be made for *all* pairs. The yardstick lengths on the $\bar{r}$ chart are 9.47, 6.70, and 4.74 mils for flights, sights, and lots, respectively.

A fallacy that must be guarded against in connection with the yardsticks is a tendency to select the greatest and the least of a series of observations and to regard them as significantly different if their heights differ by more than the length of the yardstick. This is wrong and likely to lead to far too many judgments of significant difference, because the yardstick lengths are determined on the assumption that they will be used for comparing only observations selected in a manner independent of their magnitudes. It is perfectly legitimate to use the yardstick to compare consecutive observations, or the first and last of a considerable series, though even so it must be remembered that significance will emerge fallaciously about once in 20 times. The detection of trends in quality should be facilitated by applying the yardsticks to observations a considerable distance apart. Discontinuous changes may become evident when two consecutive observations differ by an amount approximating the length of the yardstick, provided earlier observations are fairly well in accord with the first of the pair and later ones with the second. It is possible to devise exact statistical criteria for both trends and discontinuities, but workers in quality control stress the value of graphic devices.

**2.2.3. Control Chart for Mean Deflection.**—A control chart for $\bar{d}$ (Fig. 3.2) should be kept in exactly the same way as that for $\bar{r}$. However, the control limits will be farther apart in the case of the $\bar{d}$ chart and the yardsticks longer because the variance of deflection errors is greater than that of range errors. The 0.01 control limits are at 8.40 and $-2.79$ mils; the 0.05 limits are at 7.05 and $-1.44$ mils. The yardstick lengths are 11.65, 8.24, and 5.83 mils.

**2.2.4. Control Chart for Range Variance.**—For each flight the estimated variance of range errors is defined by the formula

$$s_r^2 = \frac{S(r - \bar{r})^2}{N - 1}$$

but may be found with a slight economy of labor by using the fact that

$$S(r - \bar{r})^2 = Sr^2 - \bar{r}_r Sr$$

The control chart for range variance is to be on logarithmically ruled paper (Fig. 3.3). Values of $s_r^2$ for the four flights representing each lot
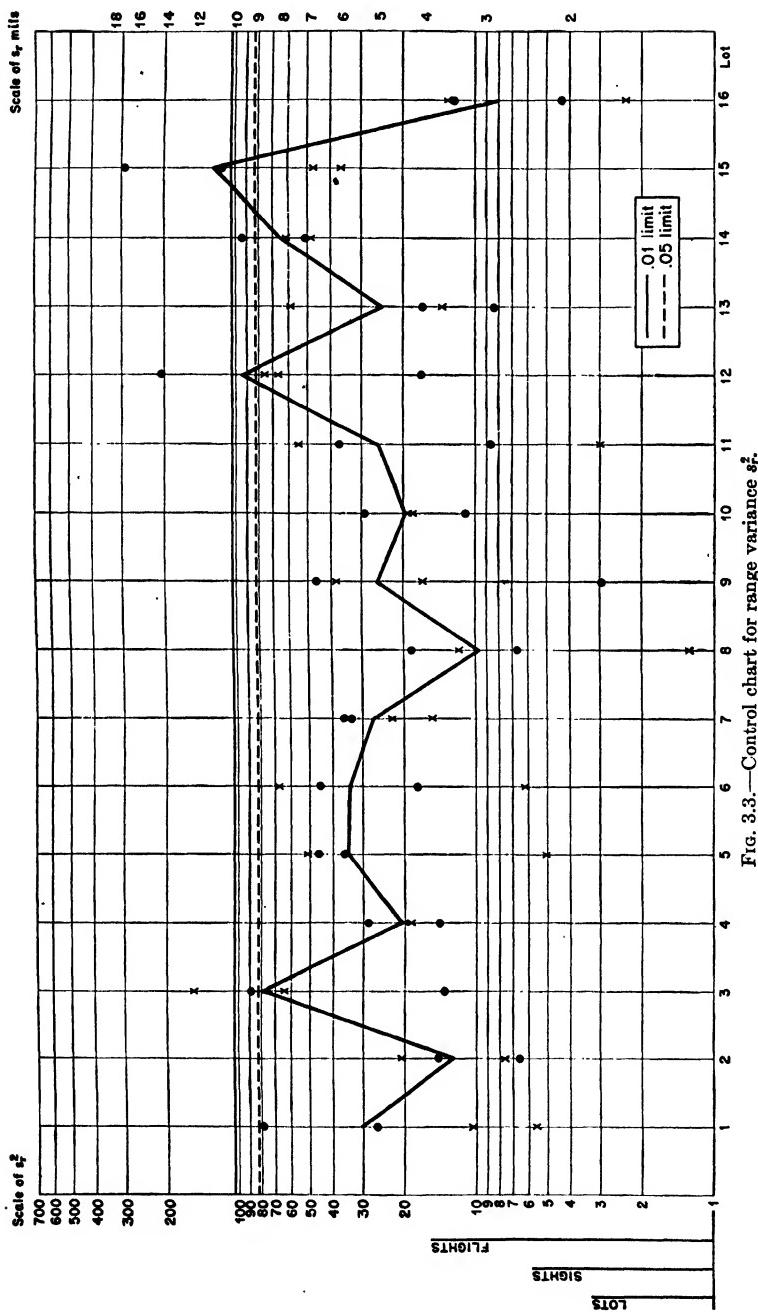
FIG. 3.3.—Control chart for range variance $s_r^2$.

of bombsights are plotted on the same vertical line, two circles for that one of the sample sights of the lot produced earlier, and two crosses for the other. The arithmetic mean of these four values of $s_r^2$ is the best estimate of range variance for this lot of bombsights. The points representing such means (which, because of the logarithmic scale, are slightly above the centers of gravity of the points representing individual values of $s_r^2$) are connected by a zigzag line. Only a single control limit is plotted in this case for each probability considered, since it is not necessary to be on guard against unusually small variances. When $N = 4$, the variance estimate for a *lot* has 12 degrees of freedom (3 for each of the 4 flights). The lot variance for range that will be exceeded by chance once in 20 times is, according to the estimate supplied by our preliminary sample, 82.87, and this is represented by the height of the control limit marked 0.05 in Fig. 3.3. The corresponding value for probability 0.01 is 104.4 (see Sec. 3.3.1 for the origin of these values).

It should particularly be observed that the denominator of each of the estimates of variance here used equals the number of degrees of freedom, not the sample number. Using the sample number, as is often done, introduces unnecessary complications into the determination of the probabilities, and if applied in the present case is likely to lead to large errors.

Yardsticks having functions similar to those of Figs. 3.1 and 3.2 are provided at the left of Fig. 3.3. The logarithmic vertical scale of Fig. 3.3 makes possible the use of these yardsticks, which are to be regarded as subject to rigid translation. The exact comparison of estimates of variance in terms of probability must be made by means of a function of their ratio, or the difference of their logarithms. This is the reason for using a logarithmic chart for $s_r^2$. In the charts for $\bar{r}$ and $\bar{d}$ an arithmetic scale was used, since the comparison between two means must be made in terms of their difference, not their ratio. The lengths of the yardsticks for flights, sights, and lots correspond, respectively, to the variance ratios 15.45, 5.82, and 3.28.

To illustrate the use of these yardsticks, suppose that $s_r^2$ should have the value 10 for one lot and 40 for the next. Both points are well within the control limits. But the ratio 4 exceeds the yardstick 3.28, as is immediately obvious when the points are charted, to such an extent that the probability is considerably less than 0.05 of the variances deviating so far from equality by chance alone. Unless some explanation connected with the testing is forthcoming, this is to be regarded as a danger signal suggesting deterioration of the sights in a manner leading to wider scatter in range. The two sights belonging to the lot showing the variance 40 should be examined for sources of excessive range variance such as loose rate knobs, and any explanation found should be indicated in the space below the control charts reserved for notes.
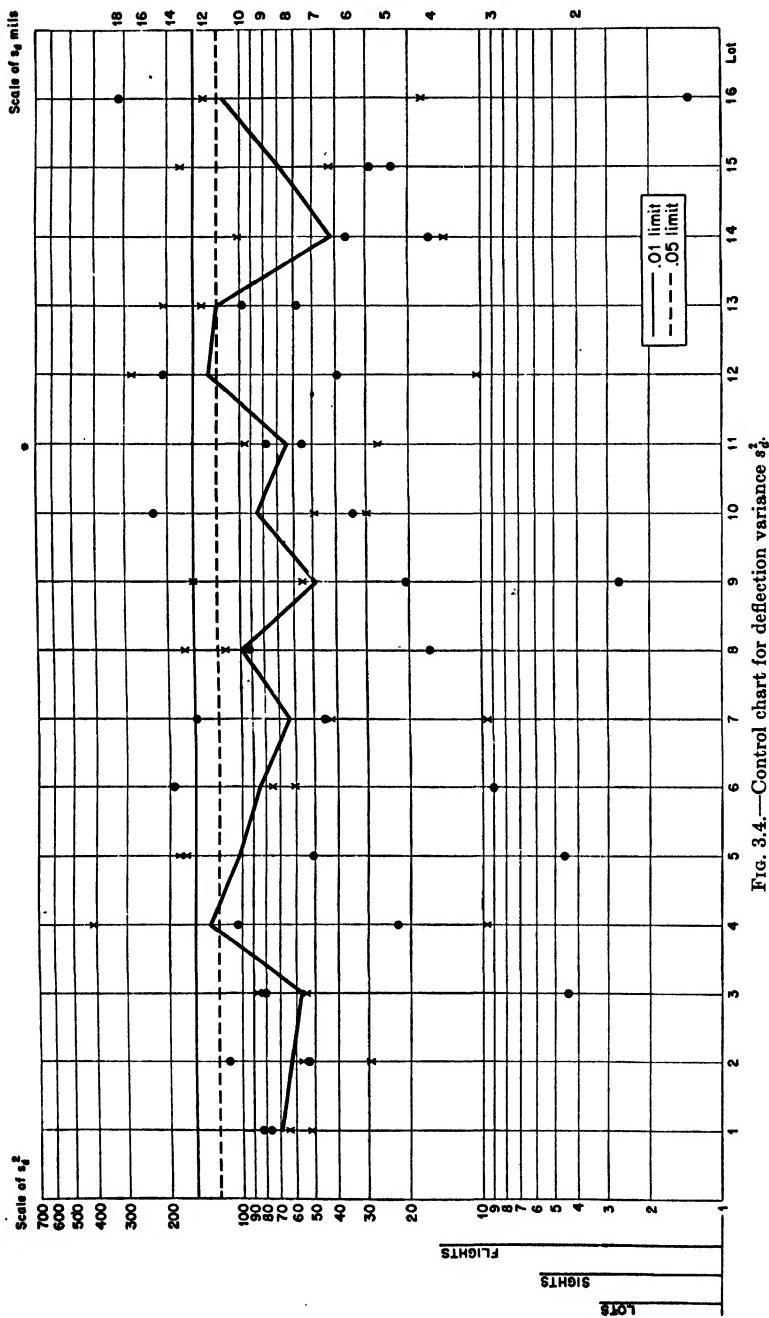
Fig. 3.4.—Control chart for deflection variance $s_d^2$.

**2.2.5. Control Chart for Deflection Variance.**—The variance of deflection errors is to be treated in exactly the same way, except that the control limits are in this case at 125.4 and 157.9. The yardstick lengths are the same as for the range-error variance chart (see Fig. 3.4).

**2.2.6. Control Charts Combining Range with Deflection.**—The indicators of quality to be used combining range and deflection errors are denoted by $T$ with the following subscripts:

> $O$ for an over-all measure
> $M$ for a measure based only on the mean point of impact (MPI)
> $D$ for a measure based on deviations from the MPI

A second subscript will also be used when needed to make the meaning clear, as follows:

> $B$ for an individual bomb
> $F$ for all the bombs dropped in a flight
> $S$ for the bombs dropped with the same sight
> $L$ for the bombs dropped with the sights representing a lot

It will frequently be convenient to omit the second subscript.

The definitions of the various $T$'s all make use of the same coefficients $a$, $b$, $c$, whose values may be taken at first, on the basis of the experience analyzed in Sec. 5, as follows:

$$a = 0.021\ 724 \qquad b = 0.001\ 331 \qquad c = 0.014\ 360$$

However, these values, and the corresponding control-chart limits, should be revised periodically as further experience accumulates. They will especially need to be revised after any change in technique of inspection has been in effect long enough to provide a sufficient basis in experience. Methods of revising them, and the other numerical quantities used in this part, are described in Sec. 3.1.3.

$T_O^2$ is defined for an individual bomb as

$$T_{OB}^2 = ar^2 + 2brd + cd^2$$

and for a flight, sight, or lot as the sum of the values of $T_{OB}^2$ for the individual bombs involved. However, it will only be in exceptional cases that $T_{OB}^2$ will actually be computed, and ordinarily it will be convenient to computed $T_{OF}^2$ by the formula

$$T_{OF}^2 = aSr^2 + 2bSrd + cSd^2$$

where, as before, $S$ stands for summation over all the bombs dropped on the flight. For a sight or a lot the value of $T_O^2$ is found by adding together its values for the several flights involved.
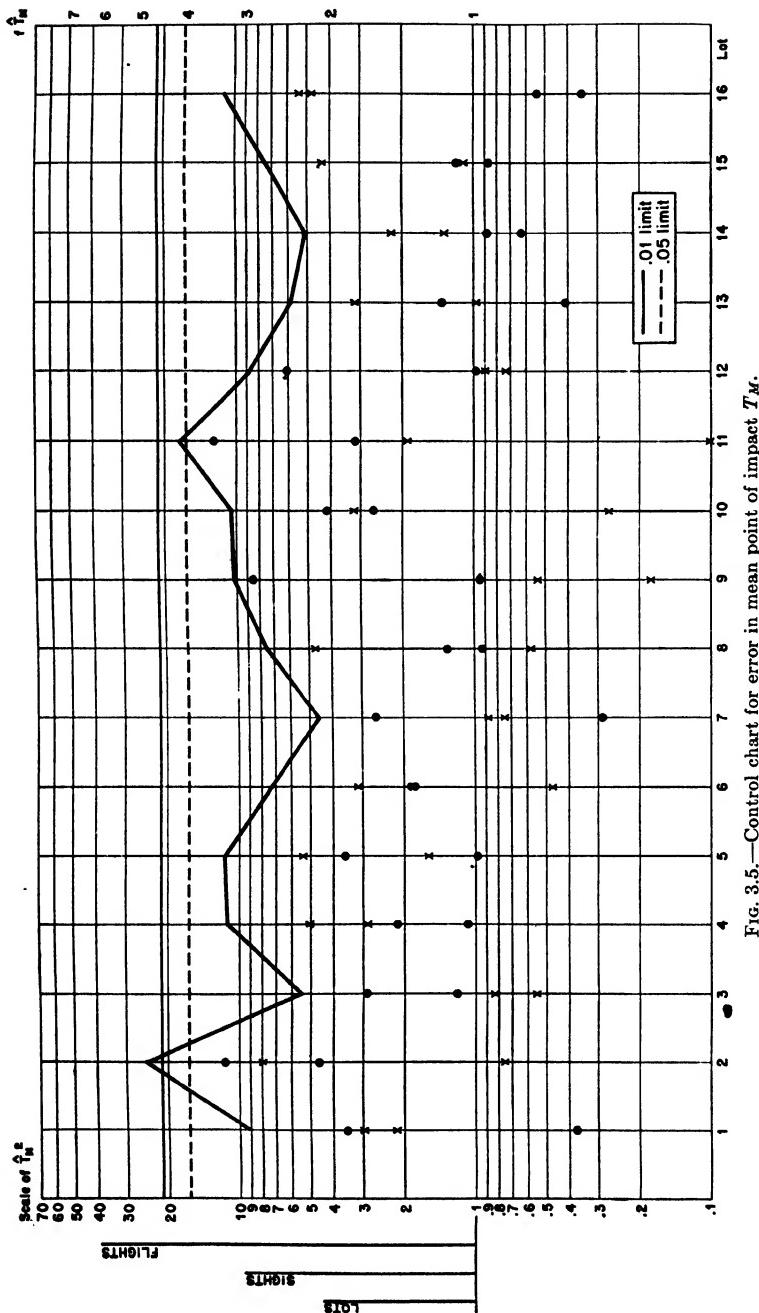
FIG. 3.5.—Control chart for error in mean point of impact $T_M$.

$T_M^2$ might be defined for a flight by the formula

$$T_{MF}^2 = N(a\bar{r}_F^2 + 2b\bar{r}_F\bar{d}_F + c\bar{d}_F^2)$$

but is more easily computed from the equivalent formula

$$T_{MF}^2 = a\bar{r}_F Sr + 2b\bar{r}_F Sd + c\bar{d}_F Sd$$

For a sight or a lot the value of $T_M^2$, like that of $T_O^2$, is found by adding up the values for the flights involved.

In every case $T_O^2$ is greater than $T_M^2$. The difference is defined as $T_D^2$. An algebraic identity shows that

$$\begin{aligned}
T_{DF}^2 &= T_{OF}^2 - T_{MF}^2 \\
&= aS(r - \bar{r}_F)^2 + 2bS(r - \bar{r}_F)(d - \bar{d}_F) + cS(d - \bar{d}_F)^2
\end{aligned}$$

while for a sight or a lot the value of $T_D^2$ is the sum of those for the flights involved. Thus in every case $T_D^2$ depends only on the deviations from the mean points of impact for the respective flights, whereas $T_M^2$ depends only on these mean points of impact.

The direct use of $T_M$ and $T_O$ will be appropriate only when biases such as those disclosed by the statistical analysis of Sec. 5 have either been eliminated or definitely attributed to the bombsight itself rather than anything in the testing process. Until such elimination is achieved or such a judgment is reached, it is better to modify $T_M$ and $T_O$ so as to make them functions of deviations, not from the target, but from the mean point of impact of the bombs of the base period. Indicating this correction by placing a caret over $T$, and using $\bar{r}'$ and $\bar{d}'$ for the coordinates of the mean point of impact of the bombs of the base period, gives

$$\hat{T}_{OF}^2 = \frac{aS(r - \bar{r}')^2 + 2bS(r - \bar{r}')(d - \bar{d}') + cS(d - \bar{d}')^2}{1 + \dfrac{1}{n'}} = \frac{T_{OF}^2 + U}{1 + \dfrac{1}{n'}}$$

$$\hat{T}_{MF}^2 = \frac{N\left[a(\bar{r}_F - \bar{r}')^2 + 2b(\bar{r}_F - \bar{r}')(\bar{d}_F - \bar{d}') + c(\bar{d}_F - \bar{d}')^2\right]}{1 + \dfrac{N}{n'}}$$

$$= \frac{T_{MF}^2 + U}{1 + \dfrac{N}{n'}}$$

where $n'$ is the number of bombs in the base period and

$$U = N(a\bar{r}'^2 + 2b\bar{r}'\bar{d}' + c\bar{d}'^2) - 2[a\bar{r}'Sr + b(\bar{r}'Sd + \bar{d}'Sr) + c\bar{d}'Sd]$$

The numerical values obtained from the 260 bombs discussed in Sec. 5 give, when $N = 4$,

$$U = 0.64917 - 0.065903Sr - 0.084129Sd$$

Fig. 3.6.—Control chart for dispersion about mean point of impact $T_D$.
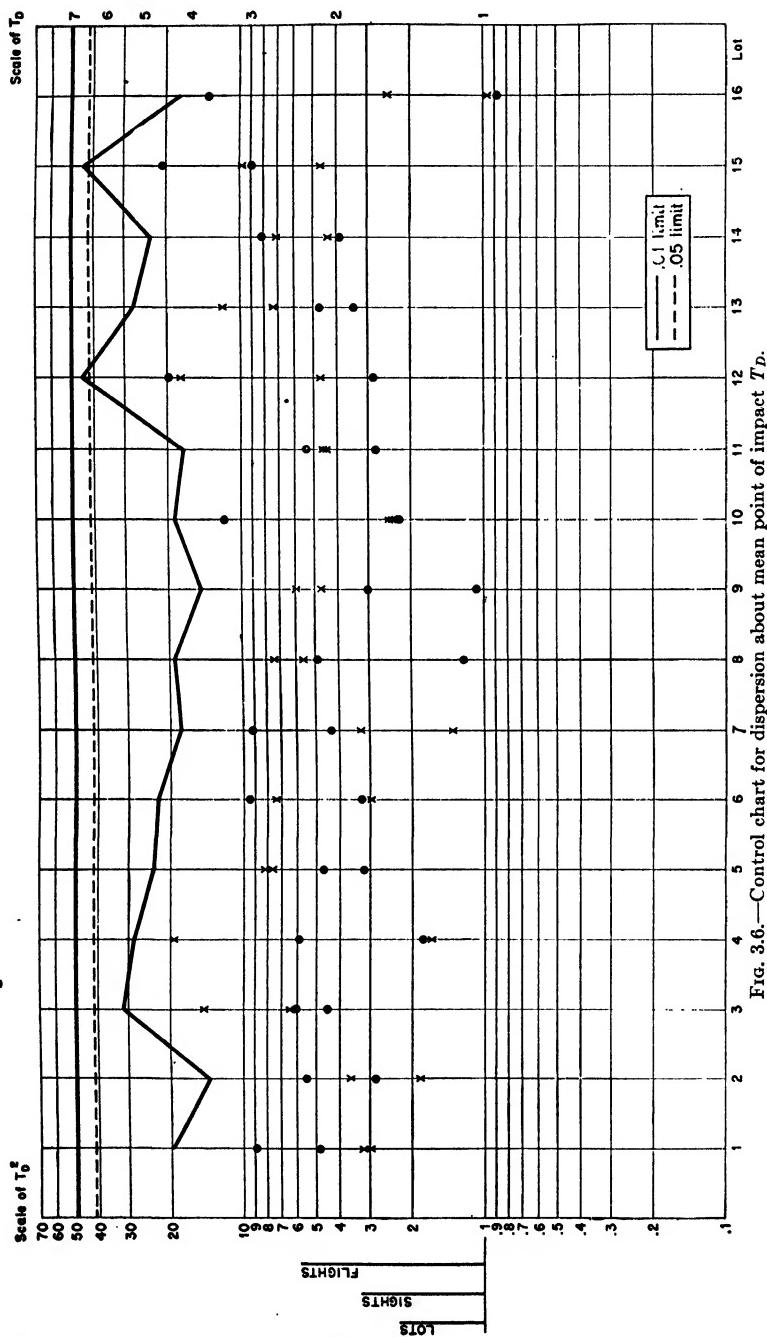
For a sight or a lot, corrected values $\hat{T}_O^2$ or $\hat{T}_M^2$ are found by adding together the corresponding corrected values for the flights involved. Values of $T_D$ are not affected by the correction.

For $\hat{T}_M^2$, $T_D^2$, and $\hat{T}_O^2$, control charts are to be maintained on arithlog paper in a similar form to those for $s_r^2$ and $s_d^2$ (see Figs. 3.5, 3.6, and 3.7). However, there is one difference in form: The zigzag line on each $T$ chart plots the *sums* of the quantities represented by circles and crosses, instead of the means as on the $s^2$ charts. This results in all the quantities plotted being of kinds customary in statistical theory, and hence easily looked up in probability tables, at least to an approximation. A further advantage of this arrangement is that it renders the appearance of the $T$ charts so unlike that of the $s$ charts that they cannot be confused.

On each of the $T$ charts there is for each of the probabilities 0.05 and 0.01 an upper but not a lower control limit. The three yardsticks supplied in each case refer to differences, on a logarithmic scale, of sums for flights, sights, and lots, respectively, with probability 0.05. The values of $T^2$ for the control limits and yardsticks are given in the following table:

| | $T_M^2$ or $\hat{T}_M^2$ | $T_D^2$ | $T_O^2$ or $\hat{T}_O^2$ |
|---|---|---|---|
| 0.05 control limit............................ | 16.3 | 41.2 | 54.4 |
| 0.01 control limit............................ | 21.2 | 49.4 | 64.0 |
| 0.05 yardsticks: | | | |
|    Flights.................................. | 39.0 | 5.8 | 4.4 |
|    Sights................................... | 9.6 | 3.3 | 2.8 |
|    Lots..................................... | 4.4 | 2.3 | 2.0 |

These values require a small correction of a purely mathematical kind based on research not yet completed.

**2.2.7. Illustration of the Use of Control Charts.**\*—A glance at Fig. 3.1 shows that all but two of the lots were well within the control limits for range MPI. Lot 9 falls slightly outside the 0.05 limit at what may be estimated as the 0.04 level. Since there is a reasonably large chance of 1 out of 15 or 16 cases exceeding the 0.05 limit, and since lot 9 shows no shortcomings on the other charts, this deviation may be disregarded.

A very different situation is presented by lot 2. As Fig. 3.1 shows, the MPI was in this case an average of nearly 6 mils short, a distance far beyond even the 0.01 control limit. Following this lot through the other control charts, we observe from Fig. 3.2 that its deflection error in MPI was very moderate, and from Fig. 3.4 that the deflection variance was exceptionally small. From Fig. 3.3 it appears that the range variance

\* This section is for illustrative purposes only and is based on data put together artificially, as explained in the footnote in Sec. 2.2.1 and in Sec. 5.
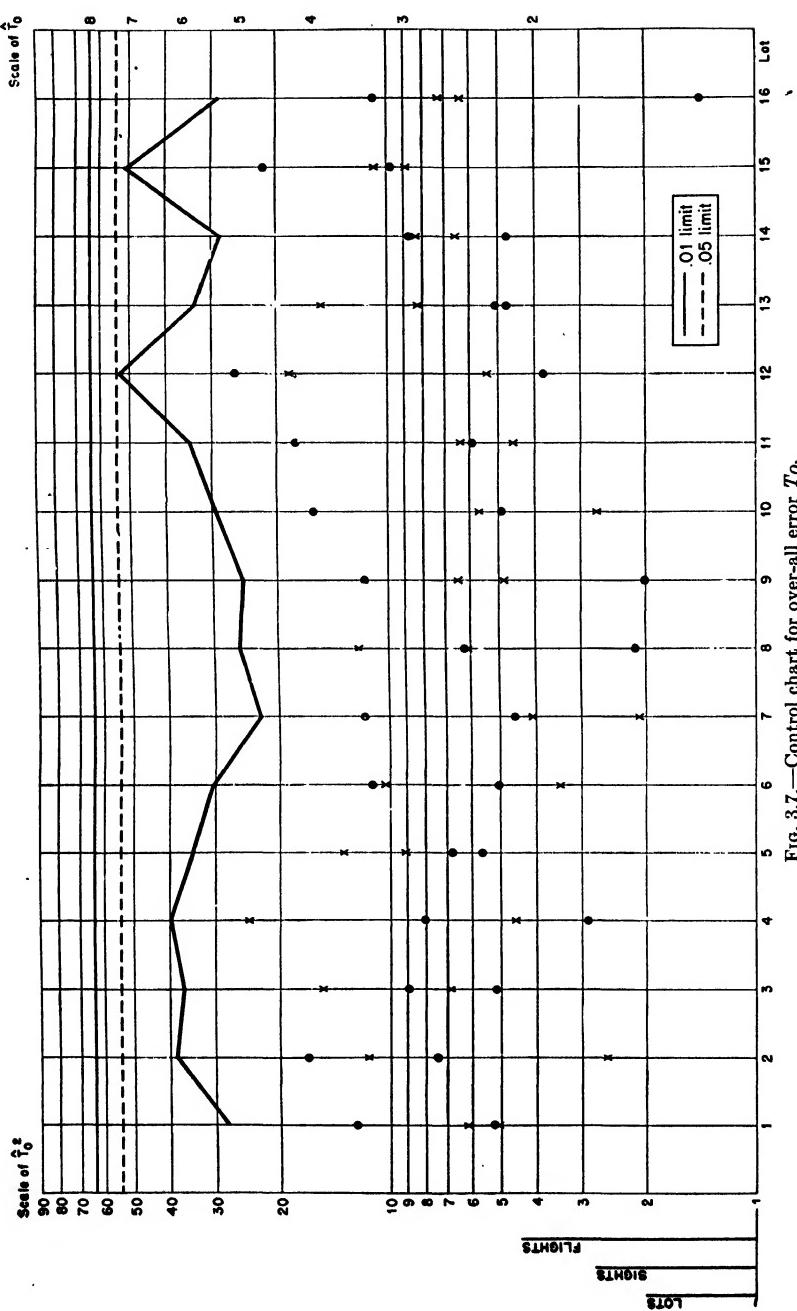
FIG. 3.7.—Control chart for over-all error $T_0$.

was far below the average.    Moreover, in Fig. 3.3 the circles and crosses for lot 2 are all remarkably close together, indicating consistently low variance, that is, high accuracy in hitting closely around a point 6 mils short of the target.    Figure 3.5 shows that the range error in MPI is too great to be concealed by averaging with the small deflection error.    Upon looking at Fig. 3.6, it is apparent that lot 2 yielded the smallest value of $T_D$ of all the 16 lots plotted, thus demonstrating a rather remarkable accuracy in hitting the wrong point.    The error in MPI and the close clustering about it offset each other sufficiently so that the bombs of lot 2 are only moderately far from the target in terms of $T_0$, so that Fig. 3.7 shows nothing remarkable.

This combination of peculiarities might be brought about by a bombardier with very quick reactions, enabling him to keep the airplane on a remarkably good course and to maintain regularity in his behavior, but causing him to release the bomb consistently too soon.    Such a man should be capable of development into an exceptionally accurate bombardier if he studied the control charts and particularly his own errors.

### 2.3. Action When Results Are Bad

If in any case the zigzag line crosses its 1 percent control limit, an investigation should be made of the cause of this anomaly, and the probable cause entered in the space provided for the purpose at the bottom of the control chart.    If the investigation includes further air testing following shop adjustment, the same seven quantities should be computed for the new air tests and their significance judged by means of the probability formulas of Sec. 3.    Entries of probable cause should indicate clearly whether the excessive bombing errors bringing about the anomaly are believed to indicate a fault in the sight or in some feature of the testing process.    Pure chance may be blamed approximately once for every hundred lots from which sample sights are tested; this allotment of 1 anomaly in 100 applies separately to each of the control charts.

Some clues as to the nature of the trouble will be provided by an intelligent reading of the control chart itself, with attention to the consistency of performance of the two sights in a lot and of the two flights with each sight, and observing whether the trouble is in range or deflection, or both, and whether it introduces a bias into the MPI or increases the dispersion about the MPI.    To facilitate such inquiries, a list of the types of defect likely to lead to excessive deviations in each of the seven control-chart quantities may be used (see that below, p. 140).

If the 0.01 control limits are exceeded repeatedly, the investigation should be extended to the factory.    Intensified shop inspection should also be applied.    In extreme cases the remaining sights in a lot represented by defective sights should be recalled or repaired.    If this happens,

the shipment of further sights prior to completion of air testing of samples should be suspended until the factory has again established a satisfactory record; and the factory should revert to the status of a new factory, which implies that 40 consecutive sights should successfully pass the air test. Exact criteria as to when this should be done will need to be formulated in the future on the basis of further experience.

When the rate of production of bombsights is to be reduced, the average showing of a factory on the control charts should be taken into consideration, in comparison with that of other factories. The average value of $T_0^2$ is probably the best criterion that can be drawn from the coordinates of the bombs dropped in air testing for deciding which factory should be eliminated.

The control chart should not, however, be looked upon primarily with a view to drastic action. Its great value, demonstrated in other industries, is in detecting trends in quality and calling attention to deterioration while it is still incipient and before it becomes serious. When this value is recognized, the attitude of manufacturers changes from hostility to appreciation. In order that these desirable results be realized, it is essential that copies of control charts be transmitted promptly to manufacturers, and that attention be called to peculiarities revealed by them. Such peculiarities include not only the crossings of the 0.01 limit that call for investigations or more serious action, but also cases of exceeding the 0.05 limit more than about once in 20 times, or of progressive drifting up toward the control limits, or of excessive variation between the performances of the same sight or of two sights from the same lot, as indicated by the yardsticks. The meaning of each of the seven control-chart indicators should be made clear to all concerned.

The yardsticks are designed to facilitate visual discernment of tendencies to drift before the control limits are reached and also to help in observing abrupt changes. A little practice and familiarity with this type of chart should make it possible to detect such changes more promptly than would otherwise be easy. Prompt detection of trends and anomalous tendencies can be of great value if the factory is notified at once, since trouble can often be found and corrected in this way before it becomes serious. It is to the interest of the government and the manufacturer alike to prevent serious trouble by early corrective action of this sort, rather than to wait for it to develop to the point at which bombsights will be rejected.

Of all the indicators, $T_0$ is the one likely to give the sharpest discrimination between cases in which sights are really defective or the bombing for any reason poor, and on the other hand cases of pure bad luck. It combines the evidence from all the others and can serve as the deciding agent in marginal cases in which the others seem to disagree on

the question whether the bombing was good or bad. In a more specialized way, $T_M$ combines the evidence of $\bar{r}$ and $\bar{d}$, while $T_D$ combines that of $s_r$ and $s_d$. It is for diagnosis of the trouble rather than for evidence of its existence that the main reliance should be placed on the charts for means and variances.

Investigations prompted by excessive control-chart deviations will be assisted by a list of possible causes under each of the seven control-chart measures. Such a list is best drawn up by persons in continuing close relation to the manufacture, testing, and use of the sights and may need modifications from time to time with changes of model and in the testing procedure. The list below is purely tentative and needs to be supplemented by further suggestions drawn from personnel concerned with manufacture, shop testing, air testing, and maintenance.

In order that such suggestions be made intelligently, it is necessary that the personnel involved have some general understanding of the meaning of each of the seven measures. The following definitions may be sufficient for most of those concerned:

$\bar{r}$    Range coordinate of mean point of impact

$\bar{d}$    Deflection coordinate of mean point of impact

$s_r$    A measure of dispersion in range

$s_d$    A measure of dispersion in deflection

$T_M$    A measure of distance from mean point of impact to target, modified in accordance with probability principles to allow for the greater deflection variance

$T_D$    A measure of dispersion about the MPI combining range and deflection errors

$T_O$    An over-all measure of errors of all kinds

It should be understood that each of the $T$ measures combines range with deflection errors, and that $T_O$ combines all the other measures into one. If a particular defect in a sight, or in the testing process, can affect only range errors it is more likely to be detected through $\bar{r}$ or $s_r$ than through any of the measures that include deflection errors. In the same way, if a particular defect can affect only deflection errors, it is more likely to be found from the $\bar{d}$ or $s_d$ chart than from any of the $T$ charts. However, any defect that causes *both* range and deflection errors may sometimes be detected through one of the combined $T$ measures when it would be impossible to distinguish it from mere chance on the basis of the charts of range or deflection errors alone. Similarly, any cause affecting both MPI and variance can sometimes be brought out clearly by $T_O$ when none of the others would make it definite.

In making up such a list, it appears better to consider the various possible sources of trouble and mark each with one or more of the seven

symbols, later rearranging to get lists in the form below, than to begin with the seven control-chart measures and try to think of all the possible causes.

The following partial and tentative list should be understood in the terms of the preceding remarks. Under each control-chart measure are listed several causes, any one of which might lead to an excessive value for the measure. Each factor is listed only under the measure or measures most likely to lead to its detection, though it may also have minor effects on other measures and will be reflected in more comprehensive measures.

$\bar{r}$

Grease on disk
Worn disk
Faulty trail plate
Release mechanism slow or fast
Bombardier's reaction slow or fast
Loose telescope cable
Faulty sight gyro
Wrong disk speed
Loose altitude knob
Slipping of scale on altitude drum
Faulty or wrongly adjusted telescope drive governor

$\bar{d}$

Faulty directional gyro
Stabilized sector off center
Pilot director scale displaced
Pilot director rheostat inaccurate
Pilot director brush worn
Crosstrail mechanism inaccurate

$s_r$

Erratic release mechanism
Erratic behavior of bombardier
Grease on disk
Roller or disk not smooth
Roller not accurately centered
Loose telescope cable
Loose range knobs
Loose altitude knob
Faulty sight gyro
Faulty telescope drive motor

$s_d$

Loose course knobs
Faulty directional gyro
Stabilized sector off center
Slow reaction by airplane
Slow reaction by pilot

$T_M$

Any combination of factors affecting $\bar{r}$ and $\bar{d}$
Wrong altitude
Wrong air speed
Faulty direction of optical axis of telescope
Sight gyro inaccurately leveled, perhaps because of inaccurate bubbles

$T_D$

Any combination of factors affecting $s_r$ and $s_d$
Loose altitude knob
Sticky bearings
Variable winds
Poor visibility
Bumpy air

$T_O$

Any combination of factors affecting two or more of the other measures
General slipshodness or lack of skill in either manufacture or testing
Variable winds
Faulty sight gyro

## 2.4. Experimentation

This plan may readily be adapted for efficient combination with experiments for determining the relative excellence of different models of sights and types of bombing, or for investigating other questions relat-

ing to bombing.   Considerable economy in the total amount of bombing required should be obtainable by a proper design for experimenting simultaneously on various types of equipment and technique while at the same time carrying out the routine test bombing with full accuracy. The amount of bombing required will be greater than for the routine testing alone or for the experiment alone, but less than the sum of the two.

To ensure valid results from such combined enterprises, it is important to verify *before* the experiment is carried out that it satisfies the criteria of the modern statistical theory of design of experiments.   It may be difficult or impossible to salvage any useful information from the results of an improperly designed experiment.   Many designs are possible for efficient combined experiments.   The selection of a particular design is best made in relation to the objects of a particular experiment.

As an illustration of a simple type of combination of experiment with routine testing, suppose that some new type of auxiliary equipment is to be examined (1) for its effect on the range component $\bar{r}$ of mean point of impact, (2) as to its effect on the standard deviation $s_r$ in range, and (3) as to its effect on general accuracy.   Instead of two flights per sight, four flights might now be made for a time with each sight.   Two of these flights are to be on one day of a 4-day period and two on another. Of the two flights with a sight on the same day, one, chosen by tossing a coin, is to carry the new equipment and the other is not.   A valid analysis of variance of range can be carried out from the results to answer question (1), and analogous procedures to answer (2) and (3) by means of $s_r$ and $T_0$, respectively, though the analysis for $T_0$ must either involve approximations or wait for mathematical research on the subject to progress a little further than the point heretofore reached.   The accuracy of determination of the differences made by the new equipment will be fully as great as if the entire enterprise were devoted wholly to this question.   From this standpoint the information relevant to the routine testing program is obtained free of charge.

For the rating of men and equipment routinely employed in the testing of bombsights, it is possible to utilize the results without any modification of design other than objective randomization in the assignments of the various flights.   This randomization should be restricted by conditions of balance and orthogonality in order to have the most accurate ratings possible and to simplify calculations, but these conditions are not essential.   One simple type of rating would be merely the average of the values of $T_0^2$ for the flights involved.   However, the use of *new types* of equipment or personnel not used in the base period from which the control limits were determined will invalidate these control limits and should not be introduced into the air testing of bombsights

without a suitable modification of design.   Similarly, altitude and other conditions attending the test bombing should be kept as nearly as possible the same as in the base period until a deliberate decision is reached that a change should be made and maintained over a considerable period.   In such an event, it will be necessary to begin over again the process of accumulating data for new control-chart limits.

### 3. FORMULAS AND CONSIDERATIONS AFFECTING REVISIONS OF THE PLAN

## 3.1. General Considerations

**3.1.1. Occasions for Revision.**—New information and changing conditions will from time to time call for revisions of the sampling plan specified in Sec. 2.   Among the changes likely to be appropriate are alterations in the size of lots, the proportion of a lot that is to be air-tested, the number of flights with each sight, the number of bombs dropped per flight, and the estimates of the variances and covariances of range and deflection errors from which control limits and yardsticks are calculated.

Another kind of change that may be developed is the use of improved types of control charts for shop inspection, together with the integration of shop-inspection procedures with air testing in a more systematic way. For such improvements the same statistical methods that are here recommended for air testing may be applied to shop testing, or to a combination of shop and air testing.   Such a development would mean that instead of the two types of measurement, range and deflection errors, it would be necessary to take account of a larger number of measures on each sight and to make appropriate combinations of the various measures.   The mathematical methods of multivariate statistical analysis are capable of extension with a little research to much more general situations involving coordinated testing at different stages. For this reason, and for the sake of other analogous applications, certain formulas will attain greater generality by being written in a form suitable for ready extension to more than two variables.   Consequently, $x_1$ and $x_2$ will sometimes be used instead of $r$ and $d$ for range and deflection errors, and subscripts will be used which take only the values 1 and 2 in the present chapter but which may be understood to take other values as well in cases in which more measurements are made.   For example, in place of the $a$, $b$, and $c$ introduced in Sec. 2.2.6, when generality is sought $l_{11}$, $l_{12}$, and $l_{22}$, respectively, will be used, so that such a formula as

$$T^2 = ar^2 + 2brd + cd^2$$

may also be written

$$T^2 = l_{11}x_1^2 + 2l_{12}x_1x_2 + l_{22}x_2^2$$

or, more compactly,

$$T^2 = \Sigma\Sigma l_{ij}x_ix_j$$

with the understanding that the first $\Sigma$ stands for summation with respect to $i$ and the second $\Sigma$ with respect to $j$.   In this chapter, the summations will be over the values 1 and 2 only, but in the extensions envisaged they may be over any set of values.   Our results will be in such forms as to be readily extended to the more general cases.

Still another type of revision of this and other sampling plans is likely to be required for sequential analysis, by which the decision as to the number of tests to be made is not fixed at the beginning of the testing but is postponed and determined by the gradually accumulating data. Considerable economies are possible by use of a well-worked-out system of sequential testing.   Methods of applying sequential analysis to the particular problem of this chapter have not yet been developed.*

**3.1.2. Numbers of Sights, Flights, and Bombs.**—In the plan suggested in Sec. 2, the proportion of sights to be sent to the proving ground was taken as one-tenth simply because this has been the proportion sent in the past and because the shop-testing facilities appear to be adequate for dealing with this proportion.   A continuing rating of the inspection at the factories in comparison with the results of the shop and air tests at the proving ground, and a consideration of the needs of the service, might lead to the decision to send either a larger or a smaller proportion of sights to the proving ground.   If it should be decided to increase the proportion of sights sent to 15 or 20 percent of the total, this could be accomplished by sending 3 or 4 sights from each lot of 20 with a minimum of disturbance of the sampling plan.   If, on the other hand, it should be decided that the inspection at the factory is more adequate than has been felt to be the case in the past, and a smaller proportion of sights is to be sent, the best way to do this would be to increase the size of the lot.   It is desirable to have the lot large enough so that, with the proportion adopted, at least two sights chosen at random from each lot shall be sent to the proving ground.   The reason for this is that there should be available a continuing measure of variation within lots, and this requires at least two sights from each lot.   Such variation is displayed by the control-chart scheme of Sec. 2 along with the measures of average quality.

If the number of sights taken from each lot should become larger, measures of this intra-lot variation could be calculated easily and accurately, but the graphic portrayal by circles and crosses might not be so clear, and vertical segments proportional to calculated measures of variation might need to be used instead.   If collateral experimentation on methods of bombing, auxiliary equipment, and so forth, is to be combined efficiently with test bombing, there is an advantage in having more

---

* See Chap. 17 for a brief account of sequential analysis.   For fuller accounts see, on applications, Statistical Research Group, *Sequential Analysis*, and on theory, Wald, "Sequential Tests."

than two sights chosen at random from each lot, as the larger numbers make possible the design of more efficient experiments. Against this consideration ·is to be set the fact that if the lots are large, as would be implied by having a large sample constituting a fixed percentage of a lot, there might be an excessive delay at the factory in the accumulation of enough sights to make a lot so that the random-sampling process could be applied. This objection to large lots would apply particularly where the rate of production of a factory is rather small. This matter is touched upon further in Sec. 4.2. The proportion of sights received at the proving ground to be air-tested has been arrived at not by calculation but somewhat arbitrarily from a general survey of the practices prevailing in the past and the opinions of those most closely in touch with various phases of the situation.

It is essential that at least two flights be made with each sight in order that a measure of variation due to the weather and chance factors may be available. The number of bombs dropped per flight has been taken to be four, because that was the most commonly used number in the period just before this study was begun. However, since six ore-filled test bombs can be dropped on a flight and since additional accuracy is gained by this practice, it may prove desirable to modify the plan to fit this or some other number $N$ of bombs dropped on a flight. Increased accuracy can also be obtained by increasing the number of flights per sight. In the light of data accumulating under the sampling plan, the question of what combination of numbers of flights per sight and of bombs per flight gives the most economical method of obtaining the requisite accuracy needs to be investigated from time to time; or (what is somewhat the same thing) what combination gives the maximum accuracy obtainable with the resources available. Such a combination can only be arrived at by a study of comparative costs in conjunction with the relation of variance between flights to variance between bombs on the same flight.

**3.1.3. Revision Based on New Experience.**—Revision of the variances and covariance of range and deflection errors may need to be made because of changes in techniques of bombing; it should eventually be made anyhow in order to have more accurate estimates of these parameters than are at present available. From a long series of reasonably homogeneous control-chart data, the variance of range $s_r^2$ may be estimated as the mean of $s_r^2$ for the recorded flights, and the variance of deflection $s_d^2$ similarly. There is a need also for an estimate, which we denote by $s_{rd}$, of the covariance of range and deflection errors. For a single flight with $N$ bombs such an estimate would be

$$\frac{S(r - \bar{r})(d - \bar{d})}{N - 1} = \frac{Srd - \bar{r}Sd}{N - 1}$$

but the average of these estimates over the long series should be adopted. (If $N$ varies among the flights, the ratio of the sum of the numerators to the sum of the denominators should be used.) It is not necessary actually to compute these sums of products if all the calculations going into the control chart have been correct and are available, since the products have already been computed in the process of obtaining values of $T^2$. Indeed (Sec. 2.2.6),

$$T^2_{DF} = aS(r - \bar{r}_F)^2 + 2bS(r - \bar{r}_F)(d - \bar{d}_F) + cS(d - \bar{d}_F)^2$$

If we sum for $m$ flights, each with $N$ bombs, and use the expressions for $s_r^2$ and $s_d^2$, we may solve and obtain as our new estimate of covariance

$$s_{rd} = \frac{1}{2b}\left(\frac{\frac{\Sigma T^2_{DF}}{m}}{N - 1} - as_r^2 - cs_d^2\right)$$

The number of degrees of freedom derived from each flight is $N - 1$, and therefore from the $m$ flights in the sample it is $m(N - 1)$. We shall put $n = m(N - 1)$. This important number, which must be used in the determination of control limits, yardstick lengths, and various tests of statistical significance, should be recorded carefully.

The revised values of the coefficients for $T^2$ are

$$a = \frac{s_d^2}{s_r^2 s_d^2 - s_{rd}^2} \qquad b = \frac{-s_{rd}}{s_r^2 s_d^2 - s_{rd}^2} \qquad c = \frac{s_r^2}{s_r^2 s_d^2 - s_{rd}^2}$$

The nature of these formulas will perhaps be clearer by reference to a more general case. If, instead of range and deflection errors, we were measuring a considerable number of features of each sight and combining these into one over-all indicator, we should use for $T^2$ the last formula on p. 142, with each coefficient $l_{ij}$ equal to the cofactor of the covariance between the $i$th and $j$th measures in the determinant of such covariances, divided by this determinant. In the formulas for $a$, $b$, and $c$ above, the common denominator is the covariance determinant, while the numerators are the cofactors of the elements of this determinant.*

After two or three years of test bombing in a fairly uniform manner, it should be possible to obtain useful measures of seasonal variation to be applied as corrections. Even without this delay, it may be possible by a special statistical study to work out corrections of the variances to allow for the weather. The effect of any such corrections upon the number $n$ of degrees of freedom should be noted.

* The reader unfamiliar with determinants can readily learn enough about them to carry out these instructions from the chapter on determinants in any book on college algebra, or from Chap. VII, "Determinants," in Crum and Schumpeter, *Rudimentary Mathematics*. In the determinant of the covariances $s_{ij}$, the elements corresponding with $i = j$ are simply variances; thus $s_{22} = s_2^2$.

**3.1.4. Assumptions Involved.**—The control limits on the charts supply a test of the hypothesis that the sights of a particular lot are as good as the sights used in the base period. The limits are so determined that, if the hypothesis is true, the 0.05 limit will in the long run be exceeded approximately once for every 20 lots, and the 0.01 limit approximately once for every 100 lots. The functions plotted on the control charts have been chosen in accordance with the findings of statistical theory as those most likely to lead to these limits being exceeded if the sights tested are in fact worse than those of the base period.

These statements are, to be sure, based on a number of assumptions, but the assumptions appear reasonable. It is assumed that the different bomb drops are statistically independent or that corrections are applied that take care of any lack of independence. If the randomization procedure prescribed in Sec. 2 is followed, and especially if suitable corrections are made for weather, this assumption should fit the facts closely as regards test bombing done under the plan. Whether true independence can be ascribed to the actual bombings used in calculating the initial control limits may be questioned, since no scheme of objective randomization was in effect at the time and since no corrections were made for weather apart from the selective effect of there being no bombing in the worst weather. It is the writer's impression, based on inspection of the procedures and records, that lack of independence in these bombings is not such as to make any great difference in the control limits. The weather factor may be more serious, since the test bombings of this base period were all made in November and December. However, the charts kept at the proving ground of earlier bombings under different conditions failed to disclose any obvious evidence of any marked seasonal variation in results.

It is assumed that the bombing errors in range and deflection have a bivariate normal distribution. This assumption has a certain a priori justification in the notion of the errors resulting from accumulation of a large number of independent errors and has also an approximate a posteriori verification in various records of bombing. There is, to be sure, a slight skewness in range errors, which can be explained by a small proportion of bombs having defects that cause them to fall short with a bias peculiar to this small group. However, the deviations from normality appear to be so small as to have no appreciable influence on the final results, so far as can be judged from general statistical considerations. Certainly there is no similarity between the recorded curves of bomb distribution and the sharp-peaked theoretical ones for which mean deviations provide efficient measures of dispersion.

No assumption is made in this chapter either that correlation of range with deflection errors exists or that it does not. The records used in

Sec. 5 indicate that little or no correlation exists, though other records consisting of compilations of huge quantities of heterogeneous material do superficially indicate high correlations. The high correlations in these cases can probably be ascribed to the mixing of records of bombings at many different places and times, with various kinds of equipment, with personnel of widely different degrees of skill, together with faulty records whose complete compilation proved impossible. Even if there were a substantial correlation, it is not clear that this would be a serious detriment if the variances of the errors were moderate. The use of $T^2$ in the plan is equally valid whether correlation exists or not.

In selecting the $T$ functions, it is assumed that no bias in any special direction is to be guarded against, but that equal precautions are to be taken against biases in all directions. Proofs of the optimum properties of $T$ have been given by Hsu and by Wald.*

It is assumed also that exact probabilities are to be represented by the control limits, free from approximations of unknown degrees of accuracy. This consideration eliminates some plausible functions, such as the mean, or mean square, of the actual distances of the bombs from the target. While these functions have an intuitive appeal and seem at first glance to represent the real value of the bombsight because closeness of bomb to target is the criterion of effectiveness in action, they are not suitable for use with proof bombing. They are inefficient even for large samples by comparison with $T_D^2$, and for small samples they suffer from the impossibility of determining exact probabilities relative to them (see Sec. 4.8).

## 3.2. Means

**3.2.1. Control Limits for Means.**—The outer control limits for the $\bar{r}$ chart are so chosen that, if the distribution of range errors is the same as in the base period, the probability is 0.01 that the mean $\bar{r}$ of the range errors found with the sights of any particular lot should be represented by a point falling outside these limits. The inner control limits are determined in the same way except that the probability is 0.05. The heights of these lines in terms of the scale of mils on the chart are given by the formula

$$\bar{r}' \pm t_e s_r' \sqrt{\frac{1}{n'} + \frac{1}{ghN}}$$

in which the following notation is used:

$\bar{r}'$ = mean of range errors in base period. (Errors "over" are considered as positive, errors "short" as negative.)

$s_r'$ = estimated standard deviation of range errors in base period,

* Hsu, "Generalized $T$"; Wald, "Tests Concerning Several Parameters."

consisting of the square root of the ratio of a sum of squares with $n$ degrees of freedom to $n$

$n'$ = number of range errors averaged in base period to obtain $\bar{r}'$

$N$ = number of bombs dropped per flight in the period for which the control chart is to be used

$g$ = number of flights per sight in the period for which the control chart is used

$h$ = number of sights per lot in the period for which the control chart is used

$\epsilon$ = 0.01 for outer control limits, 0.05 for inner limits

$t_\epsilon$ = value of ratio of deviation in a mean to its estimated standard error (with $n$ degrees of freedom) exceeded in absolute value with probability $\epsilon$

With the additional notation

$\bar{d}'$ = mean of deflection errors in base period (errors to the right considered positive, to the left as negative)

$s'_d$ = estimated standard deviation of deflection errors in base period

we have the parallel formula for control limits for $\bar{d}_L$

$$\bar{d}' \pm t_\epsilon s'_d \sqrt{\frac{1}{n'} + \frac{1}{ghN}}$$

The base period considered in Sec. 5 yields the following numerical values:

$$\bar{r}' = 1.3450 \text{ mils} \qquad \bar{d}' = 2.8046 \text{ mils}$$
$$s'_r = 6.8040 \text{ mils} \qquad s'_d = 8.3686 \text{ mils}$$
$$n' = 260 \qquad n = 256$$

According to the specifications in Sec. 2.1, $g = h = 2$ and $N = 4$. Values of $t_\epsilon$ can be found in tables of the Student distribution.* However, these tables do not extend to the value $n = 256$. For such large values $t_\epsilon$ may be obtained as the first two or three terms of the asymptotic series

$$t_\epsilon = x \left[ 1 + \frac{x^2 + 1}{4n} + \frac{(x^2 + 3)(5x^2 + 1)}{96n^2} + \cdots \right]$$

where $x$ is the unit normal deviate exceeded in absolute value with probability $\epsilon$.† We thus find for $n = 256$

* See FISHER and YATES, *Statistical Tables*, Table III. This shows $t_\epsilon$ to 3 decimal places for $\epsilon = 0.001, 0.01, 0.02, 0.05,$ and 0.1 to 0.9 by steps of 0.1, for $n = 1$ to 30 by steps of 1, 40, 60, 120, and $\infty$. Also, see MERRINGTON, "Percentage Points of the $t$-Distribution." This shows $t_\epsilon$ to 5 significant figures for $\epsilon = 0.005, 0.01, 0.025, 0.05, 0.10, 0.25,$ and 0.50, for the same values of $n$ used by Fisher and Yates.

† Values of $x$ may be found from Table 1.1 in Chap. 1, provided $\epsilon/2$ is used in

$$t_{0.01} = 2.59514 \qquad t_{0.05} = 1.96925$$

Substituting these values gives control limits for $\bar{r}_L$ as follows:

$$\text{0.01 limits:} \quad 1.3450 \pm 4.5481 = 5.8931 \text{ and } -3.2031$$
$$\text{0.05 limits:} \quad 1.3450 \pm 3.4512 = 4.7962 \text{ and } -2.1062$$

For $\bar{d}_L$, in the same way, we find

$$\text{0.01 limits:} \quad 2.8046 \pm 5.5940 = 8.3986 \text{ and } -2.7894$$
$$\text{0.05 limits:} \quad 2.8046 \pm 4.2448 = 7.0494 \text{ and } -1.4402$$

**3.2.2. Asymmetry of Control Limits and Bias in Base Period.**—It may seem strange that the control limits are not symmetrical with respect to the zero line on the $\bar{r}$ and $\bar{d}$ charts, since sights whose use gives mean points of impact close to the target seem more desirable than those yielding more remote MPI's. A like question arises in connection with $T_M$ and $T_O$. However, it should be emphasized that our comparisons are not with a theoretical ideal but with the actual performance of a base period in which the testing procedure is presumed to be comparable with that in the period for which the control chart is to be used. The material on which our records are based shows biases in both range and deflection errors in excess of any that could reasonably be ascribed to chance. Whether these biases resided in the sights, in the habits of the test bombardiers or pilots, in some other feature of the test bombing, or in the observation of the results we do not know. The sights used were accepted. If the bias was the fault of the sights, this fact should be ascertained by an *ad hoc* inquiry and the manufacturing design changed accordingly. If the bias was a result of something in the test bombing other than the sights, and if this condition continues in the period for which the control charts are to be used, it is necessary that the control limits be asymmetrical with respect to the zero line in order to avoid wholesale false alarms regarding sights that are really quite satisfactory but with which the test bombs fall moderately farther over and to the right than the average. If the proof-bombing procedure is such as to create the observed bias, a sight should be labeled unsatisfactory much more readily for bomb errors to the left and short than for errors of the same magnitude in the direction in which the bomb errors are usually made. Use of the asymmetrical limits is equivalent to applying a correction to each bomb coordinate to compensate for the bias observed in the base period.

---

entering the table. The value of $x$ for $\epsilon = 0.01$ is 2.57582930, and for $\epsilon = 0.05$ it is 1.95996398.

    The series for $t_\epsilon$ in terms of $x$ is equivalent to a series for $x$ in terms of $t_\epsilon$ given by Hotelling and Frankel, "Transformation of Statistics," and by Goldberg and Levine, "$t$ and $\chi^2$"; the first two terms are also given by Peiser, "Significance Levels."

If a change in bombing procedures, or in methods of observing the positions at which the bombs fall, results in elimination or alteration of the observed· bias, the control limits should be changed accordingly· Until this is done they should be based on experience.

If an investigation leads to the conclusion that freedom from bias has been achieved in the test bombing and is likely to continue, the control limits for $\bar{r}$ and $\bar{d}$ should be symmetrical about zero, at the same distances as from the central lines previously indicated. At the same time, $\hat{T}_M$ should be replaced by $T_M$, and $\hat{T}_O$ by $T_O$, both as to calculated values and in the labeling of the control chart.

Other considerations involving the fallibility of the base-period sample arise in connection with the effect on the distance between upper and lower control limits of the number of bombs in the base period. The reciprocal of this number $n'$ is one of the terms under the radical sign in the formulas for the control limits, so that if the control limits are based on a period in which very few bombs were dropped, the limits will be relatively far apart. In such a period $t_\epsilon$ will also be slightly larger than if numerous bombs were dropped and will push the control limits still farther apart. This fact is a consequence of our ability to make clear comparisons only with past observations. If the preliminary sample with which we compare a new sample is small, the possibility of establishing a difference when one really exists is diminished.

The effect of $n'$ on the control limits is rather slight for numbers of any such orders of magnitude as we have been considering. There seems no reason here why the control limits should not be revised after every 200 or 300 bombs, if it is thought that fundamental conditions have changed in such an interval. A more serious objection to such frequent revision for the present is the lack of exact probabilities when $n$ is small.

**3.2.3. Yardsticks for Differences of Means.**—Unlike the control limits, these yardsticks are unaffected by the bias in the MPI's of the base period. The lengths of the yardsticks for means of flights, sights, and lots are, respectively,

$$t_{0.05} \sqrt{\frac{2}{N}}\, s' \qquad t_{0.05} \sqrt{\frac{2}{gN}}\, s' \qquad t_{0.05} \sqrt{\frac{2}{ghN}}\, s'$$

where $s'$ is to be replaced by $s'_r$ for the range yardsticks and by $s'_d$ for the deflection yardsticks, and the value of $t_{0.05}$ is to be determined with reference to the value of $n$ underlying the estimates of variance in the base period. For $g = h = 2$, $N = 4$, and the same values of $s'_r$ and $s'_d$ used previously on the basis of experience, these expressions take the following values, which are the lengths of the yardsticks shown in Figs. 3.1 and 3.2:

| | Flights | Sights | Lots |
|---|---|---|---|
| Range.................................. | 9.4741 | 6.6992 | 4.7370 |
| Deflection........  ......................... | 11.6527 | 8.2397 | 5.8263 |

## 3.3. Variances

**3.3.1. Control Limits for Variances.**—The control limits on the charts for $s_r^2$ and $s_d^2$ are fixed in accordance with the distribution of the ratio of two independent estimates of the same variance. This ratio is usually denoted by $F$, since R. A. Fisher was the first to work with the distribution and to show its extensive usefulness in statistical work of many kinds.*

To obtain the 0.01 control limit for $s_r^2$, we put $s_r^2/s_r'^2 = F$, and obtain from the table of 1 percent points the value of $F$ corresponding to

$$n_1 = gh(N - 1) \quad \text{and} \quad n_2 = n$$

From this value of $F$ and the value $s_r'^2$ found from the base period we obtain $s_r^2$, the upper control limit. According to the plan specified in Sec. 2.1, $g = h = 2$ and $N = 4$, so that $n_1 = 12$. For the base period of Sec. 5, $n = 256$ and $s_r'^2 = 46.295$. From the tables it is found that, for $n_1 = 12$ and $n_2 = 256$, the 1 percent point of $F$ is approximately 2.2544. Multiplying this by 46.295 gives the value 104.4 as the 1 percent control limit for $s_r^2$.

The numerical values for the control limits on the $s^2$ scale of Figs. 3.3 and 3.4 are found in this way and given below:

| | 0.01 limit | 0.05 limit |
|---|---|---|
| Range.................................. | 104.4 | 82.9 |
| Deflection.............................. | 157.9 | 125.4 |

These control limits take full account of the fact that the variances derived from the preliminary sample are not exactly those of the population but are based on 256 degrees of freedom instead of an infinite number. If the same variance estimates had been obtained from a preliminary sample of, say, 1,000,000 degrees of freedom, the control limits for range variance would have been reduced to 101.2 and 81.1, with proportionate reductions in the case of deflection.

* For a table of $F$, see Fisher and Yates, *Statistical Tables*, Table V, where $F$ is called "the variance ratio." This table contains a footnote which, since it may be misinterpreted, should be disregarded in following the instructions in this chapter. A more extensive table of $F$ is given by Merrington and Thompson, "Percentage Points of the Inverted Beta ($F$) Distribution."

**3.3.2. Yardsticks for Variances.**—The yardsticks on the variance charts have no relation whatever to the base period but depend only on the numbers of degrees of freedom in the current variance estimates that are to be compared. Their lengths are the same for the charts of range variance and of deflection variance. The vertical distance between two plotted points has probability 0.05 of exceeding the length of the appropriate yardstick if the variances plotted are obtained from samples from the same hypothetical infinite population. If $F$ is the ratio of two such sample variances with equal numbers of degrees of freedom, $\log F$ has a symmetrical distribution extending from minus to plus infinity. The value of $\log F$ corresponding to the length of the yardstick is such that its probability of being exceeded *in absolute value* is 0.05. Hence, the probability of a random value being actually greater than this is 0.025, and the probability of its being less than the negative of this yardstick value is also 0.025. The yardstick length would thus correspond to the $2\frac{1}{2}$ percent point of the $F$ distribution.*

For each flight the number of degrees of freedom for the differences among the range coordinates of the $N$ bombs is $N - 1$; this is equal to the number of degrees of freedom of the deviations of these $N$ numbers from their mean. The variance estimates for sights and for flights are obtained by adding the sums of squares of deviations from the respective flight means and dividing by the sum of the numbers of degrees of freedom. Thus the number of degrees of freedom in the variance estimate for a sight is $g(N - 1)$, while for a lot it is $gh(N - 1)$.

For the specified values $g = h = 2$, $N = 4$ the numbers of degrees of freedom for flights, sights, and lots, respectively, are therefore 3, 6, and 12. The critical values of $F$ in these three cases are 15.439, 5.8197, and 3.2773. On the logarithmic scale of $s^2$ at the left of Fig. 3.3 the heights from 1 to these three numbers are the lengths of the yardsticks. On Fig. 3.4 the lengths are the same.

## 3.4. Comprehensive Measures

**3.4.1. Control Limits for the $T$'s.**—When the variance estimates for the base period utilize a large number $n$ of degrees of freedom, such as the value 256 here used, the distribution of $T^2$ is approximately the same as the $\chi^2$ distribution that has been tabulated and widely used for a variety of statistical purposes.† This distribution depends on a number of degrees of freedom $f$, which may be defined in the following manner. For each bomb the measure of discrepancy from the target

* The $2\frac{1}{2}$ percent points of $F$ are not given by Fisher and Yates but are given by Merrington and Thompson.

† For tables of the $\chi^2$ distribution, see Fisher and Yates, *Statistical Tables*, Table IV; or Thompson, "Percentage Points of the $\chi^2$ Distribution."

$$T_{OB}^2 = ar^2 + 2brd + cd^2$$

has (approximately for large $n$) the $\chi^2$ distribution with 2 degrees of freedom, provided the bomb is dropped under essentially the same conditions as the bombs in the base period, and provided the probability distribution is actually centered at the target without bias. The sum of any number of independent values of $\chi^2$ has the $\chi^2$ distribution with a number of degrees of freedom equal to the sum of the numbers for the several values of $\chi^2$ added. Thus $T_{OF}^2$, the sum of the $N$ values of $T_{OB}^2$ for the $N$ bombs dropped in a flight, has approximately the $\chi^2$ distribution with $2N$ degrees of freedom. Since $g$ flights are made with each sight, $T_{OS}^2$ has $2gN$ degrees of freedom, and since $h$ flights of each lot are tested, $T_{OL}^2$ has $2ghN$ degrees of freedom. Since in the specific plan of Sec. 2.1, $g = h = 2$ and $N = 4$, control limits for $T_{OL}^2$ may be obtained approximately from the 0.01 and 0.05 probability levels for $\chi^2$ with 32 degrees of freedom.

Under this same condition of absence of bias, $\bar{r}_F \sqrt{N}$ and $\bar{d}_F \sqrt{N}$ have the same joint distribution as $r$ and $d$. Consequently

$$T_{MF}^2 = N(a\bar{r}^2 + 2b\bar{r}\bar{d} + c\bar{d}^2)$$

has, with the same approximation as before, the $\chi^2$ distribution with 2 degrees of freedom. For the $gh$ flights involved in testing a sample lot, the sum $T_{ML}^2$ of these quantities has approximately the $\chi^2$ distribution with $2gh$ degrees of freedom. For the plan of Sec. 2.1 this means 8 degrees of freedom for $T_{ML}^2$.

The distribution of a sum of squares of deviations from a mean is independent of the distribution of the mean itself when, as is assumed throughout, the fundamental distribution is of the normal form. In like manner the sum of the products of the deviations of $r$ and $d$ from their respective means is independent of these means in the probability sense. It follows that

$$T_{DF}^2 = aS(r - \bar{r}_F)^2 + 2bS(r - \bar{r}_F)(d - \bar{d}_F) + cS(d - \bar{d}_F)^2$$

is independent of $T_{MF}^2$. Since

$$T_{DF}^2 + T_{MF}^2 = T_{OF}^2$$

and since $T_{MF}$ has 2 degrees of freedom while $T_{OF}$ has $2N$, it follows that $T_{DF}^2$ has approximately the $\chi^2$ distribution with $2(N - 1)$ degrees of freedom. Since $T_{DL}^2$ is the sum of $gh$ independent values of $T_{DF}^2$, the number of degrees of freedom for $T_{DL}$ is $2gh(N - 1)$, a number which for the plan of Sec. 2.1 equals 24.

It is assumed in the foregoing paragraphs that $r$ and $d$ are free from bias. If there is a bias, the same in the new as in the old bombings,

$r - \bar{r}'$ and $d - \bar{d}'$ will be free of bias but will have variances $1 + 1/n'$ times those of $r$ and $d$, respectively. (Here $n'$ is the number of bombs in the old sample, equal to 260 for the base period considered.) Hence, if we replace $r$ and $d$, respectively, by

$$\frac{r - \bar{r}'}{\sqrt{1 + \frac{1}{n'}}} \quad \text{and} \quad \frac{d - \bar{d}'}{\sqrt{1 + \frac{1}{n'}}}$$

in $T_{OB}^2$, the resulting statistic

$$\hat{T}_{OB}^2 = \frac{a(r - \bar{r}')^2 + 2b(r - \bar{r}')(d - \bar{d}') + c(d - \bar{d}')^2}{1 + \frac{1}{n'}}$$

has the distribution that $T_{OB}^2$ would have in the unbiased case. The sum $\hat{T}_{OL}^2$ of $Ngh$ values of $\hat{T}_{OB}^2$ therefore has the same approximate $\chi^2$ distribution with $2Ngh$ degrees of freedom as $T_{OL}^2$.

The mean range $\bar{r}_F$ for the $N$ bombs of a flight differs from the mean range $\bar{r}'$ for the $n'$ bombs of the base period by an amount whose variance is $1/N + 1/n'$ times the variance of range in general. The like is true of the difference $\bar{d}_F - \bar{d}'$. Consequently

$$\hat{T}_{MF}^2 = \frac{a(\bar{r}_F - \bar{r}')^2 + 2b(\bar{r}_F - \bar{r}')(\bar{d}_F - \bar{d}') + c(\bar{d}_F - \bar{d}')^2}{\frac{1}{N} + \frac{1}{n'}}$$

has the same approximate $\chi^2$ distribution as $T_{MF}^2$ with $2gh$ degrees of freedom.

These approximate distributions require correction for the limited number $n$ of degrees of freedom of the sample in the base period. If a base period with a value of $n$ less than, say, 100 were taken, the correction would be substantial. As the value of $n$ increases, the exact $\chi^2$ distribution is approached as a limit.

No study of the exact distributions of the various $T$ statistics here used has been published except for the simplest case. The probability that $T_{OB}$ should exceed any assigned value $T$ is known to be[*]

$$\left(1 + \frac{T^2}{n}\right)^{-\frac{n-1}{2}}$$

This is the special case for $p = 2$ of the distribution of the generalized Student ratio given by

$$T^2 = \Sigma\Sigma l_{ij}x_i x_j$$

where $l_{ij}$ is the element in the $i$th row and $j$th column of a $p$-rowed symmetric matrix inverse to the matrix of sample covariances $s_{ij}$ between $x_i$

[*] See HOTELLING, "Generalization of Student's Ratio," formula (26).

and $x_i$ found in a former sample with $n$ degrees of freedom for each of the $p$ variates. The element of probability for this generalized Student ratio, namely,

$$\frac{2\Gamma\left(\dfrac{n+1}{2}\right)}{\Gamma\left(\dfrac{p}{2}\right)\Gamma\left(\dfrac{n-p+1}{2}\right)n^{p/2}\left(1+\dfrac{T^2}{n}\right)^{(n+1)/2}}\frac{T^{p-1}dT}{}$$

approaches that for $\chi$ with $p$ degrees of freedom when $n$ increases. The probability integral of this distribution may be evaluated in various ways, one of which is to use the asymptotic expansion*

$$\chi^2 \sim T^2\left[1 - \frac{p+T^2}{2n} + \frac{4-p^2+(2+5p)T^2+8T^4}{24n^2} \cdots \right]$$

in conjunction with the $\chi^2$ distribution. From it may be obtained the asymptotic expansion for $T^2$ in terms of $\chi^2$, namely,

$$T^2 \sim \chi^2\left[1 + \frac{p+\chi^2}{2n} + \frac{7p^2-4+(13p-2)\chi^2+4\chi^4}{24n^2} \cdots \right]$$

This formula is particularly suitable for determining the values of $T^2$ corresponding to any particular probability level from those already tabulated for $\chi^2$.

Probabilities connected with the $T$ distribution, or values which will be exceeded with assigned probabilities, can also be evaluated by means of the transformation

$$T^2 = \frac{npF}{n-p+1}$$

where $F$ has the variance ratio distribution, of which tables are available, with $n_1 = p$ and $n_2 = n - p + 1$ degrees of freedom. This method is particularly suitable for small values of $n$.

This known $T$ distribution may be used temporarily as an approximation for the distributions of the $T$ statistics used in the plan presented in this chapter. It has been so used in obtaining the control limits listed in Sec. 2.2.6, which are shown in Figs. 3.5, 3.6, and 3.7. For this approximation the $p$ of the foregoing formulas is replaced by the number of degrees of freedom $f$, which in Sec. 2 has the values 8, 24, and 32 for $T_{ML}$, $T_{DL}$, and $T_{OL}$, respectively. In general, $f$ equals $2gh$, $2gh(N-1)$, and $2ghN$ for these several cases.†

* See Hotelling and Frankel, "Transformation of Statistics," p. 96.

† An article on the mathematical theory of this chapter, which is expected to appear in the *Annals of Mathematical Statistics*, will include the exact distribution of these $T$'s.

**3.4.2. Yardsticks for the $T$'s.**—The values of $T^2$ obtained from different flights are not exactly independent in the probability sense, since they involve the same values of $a$, $b$, and $c$, which are subject to the sampling errors of the base period, which will be small if $n$ is large. For this reason it appears that the several values of $T^2$ will have positive correlations, but that these will be very small when $n$ is as large as in the sample here considered. Apparently no great error will be made by assuming independence.

Pending the development of a more exact treatment, we may obtain yardsticks for the $T$ charts by assuming independence and attributing to each $T$ the $\chi$ distribution. The ratio of two independent values of $T^2$, each with $f$ degrees of freedom, has thus approximately the distribution of the ratio of two independent values of $\chi^2$, that is, $F$, with $n_1 = n_2 = f$. Since yardsticks are desired with a probability 0.05 of being exceeded in length on the logarithmic scale, the tables of the $2\frac{1}{2}$ percent points of $F$ may be used. The values of $f$ for single flights are $2N$, 2, and $2(N-1)$ for $T_O$, $T_M$, and $T_D$, respectively. These values of $f$ are to be multiplied by $g$ for sights and by $gh$ for lots.

## 4. BACKGROUND AND REASONS FOR THE PLAN

### 4.1. Objectives of Sampling Inspection

Apart from special experimental and exploratory purposes, two principal objects can be gained efficiently by means of routine inspection and testing of small samples, and a third object can at the same time be attained in a somewhat less adequate fashion. From small samples it is possible

(1) To judge efficiently the general level of quality of the product of a factory or a part of a factory over a period, in comparison with the output of other productive units, and of the same unit at other times.

(2) To detect promptly irregularities, trends in quality, and deviations from customary levels.

(3) To select and reject individual bad articles, or bad lots. This function is often considered the primary one. However, the efficiency of small samples for this purpose, when the proportion of defectives that can be tolerated is small, is not so satisfactory as is sometimes imagined. This has beeen brought out in a recent article, "Rational Sentencing," by Dr. E. H. Sealy.

On the basis of the small samples (5 to 10 per lot of 2,000) which it is possible to fire at ammunition proof, it is quite impossible to distinguish between the qualities of individual lots, or to sort out individual bad lots from the bulk in any but the grossest cases. Proof specifications of the type which state that a lot will be accepted if its proof is clear and rejected or reproved if certain defects appear at proof have very little effect upon the quality of the bulk being inspected.

This has often been demonstrated in cases where the test is nondestructive by 100 percent inspecting the bulk both before and after proof. The percentages defective found at 100 percent inspection of the accepted and rejected lots are almost invariably found to be nearly equal and both very little different from the percentage of defectives in the original bulk presented for proof. In other words, the quality of the product passed into the service under such a sentencing system is practically that which happens to be made at the factory. . . .

It is, then, quite impracticable to attempt to discriminate between individual lots by means of a small sample proof. What we *can* get, however, from such a proof is a knowledge of the average, over-all quality of the product of any given factory, for the accumulation of many small sample proofs amounts in time to bulk of evidence which there is no gainsaying. . . .

Broadly speaking, then, the object of rational sentencing is not to sort out individual good and bad lots but to estimate qualities assumed to be reasonably constant. If we estimate that, over a long period, a firm produces consistently a product no better than 3 percent defective, then if we need this firm's production, we must perforce accept it as a whole along with its 3 per cent of defectives, for this is the best this firm can do under its present state of knowledge and equipment. In such a case there will be, on the average, about 60 defectives in each lot of 2,000, and clearly, by chance, when picking a sample at random from a lot, one or more of these 60 defectives will from time to time be found. But it is irrational to suppose that the lots, none of whose 60 defectives chanced to be chosen in the proof sample, are any better than the others.

Dr. Sealy has elsewhere* shown the great difficulty of devising a practical small-sample scheme that will reduce to reasonable dimensions both the producer's risk of having good lots rejected and the consumer's risk of having bad lots accepted, when the definition of goodness is that the proportion defective in a lot shall be not more than about 2 percent. His work is, however, confined to cases in which the result of the inspection of each article is merely to classify it as defective or passable. The crudeness of this type of inspection in comparison with quantitative measurements is partly responsible for the difficulties he emphasizes. With the relatively refined statistical methods recommended in this chapter for quantitative measures of bombsight quality, the chances of detecting a bad lot are much greater than by the simple procedures whose limitations he points out. The fact remains, however, that too much should not be expected of small samples from the standpoint of catching infrequent defects.

One way to meet this difficulty is to increase the size of the sample by lumping together successive lots, while keeping the same proportion of inspection, so as to have larger samples. This procedure, however, introduces new and formidable difficulties. It is objectionable to keep lots of thousands on hand for the long periods required for selection and

* "Acceptance Sampling."

testing of samples; and, more fundamentally, production conditions are likely to change considerably during the accumulation and testing processes for such samples, so that the homogeneity necessary for accurate inferences is lost.

### 4.2. Reasons for Random Selection and for Lot Size 20

It is evident that the detection of individual defective bombsights, or small lots of bombsights, needs to be done primarily by tests in the shop rather than in the air. The main reliance for this purpose must be on the painstaking, detailed inspection of every sight made by the government inspectors at the factory. The inspection at the proving ground of 10 percent of the sights can be only a partial check on the inspectors at the factories, incapable of reliability in detection of bad small lots or individual sights in any but the grossest cases.

The real functions of testing and inspection at the proving ground must be taken primarily to be the two set forth at the beginning of the preceding section, namely, (1) the continuing appraisal of general level of quality, and (2) the detection of trends, excessive irregularity, and differences between sights produced at about the same time in the same factory.

To carry out these functions it is essential to have at the proving ground sights chosen at random. Selection of those apparently the worst, as has been done in the past, introduces not only a bias, which could be compensated by calculation after the accumulation of experience with sufficient numbers chosen at random, but the much more serious difficulty of excessive variation. To carry out accurately the two leading functions of sampling inspection on the basis of samples thus selected is practically impossible. The worst, or the least, or the greatest, with respect to any characteristic, in samples of 10 or more has a distribution far removed from that of the general population from which the samples were drawn, not only in average level, but also in variability from sample to sample and in the form of the distribution. It is these considerations that lead to the recommendation to select 10 percent of the output at random, in place of taking the worst 1 in 10.

We do not recommend taking 1 in each successive lot of 10 at random, but rather taking 2 at random from each of successive lots of 20. This might be replaced by taking 3 at random from each of successive lots of 30, or similarly with large numbers. However, the use of these larger lots would mean delay while they were being accumulated and would increase the trouble with fractions of lots remaining as tag ends with every change of model. An even more cogent objection to large lots is that the larger the lot the greater is the chance of some change in the conditions of production while it is being accumulated. Such a change means

heterogeneity within the lot, which weakens the inferences from it, and means that the existence of the change is not likely to be detected so soon as if the lots were smaller. It is particularly important that deteriorations in quality be discovered promptly, in order that further defective sights may be prevented before a great number of them accumulates. These considerations lead to a choice of size of lot as small as is consistent with having a plurality of individual sights chosen at random from it.

The use of 10 percent as the proportion of sights to be sent to the proving ground is neither defended nor opposed here but is adopted as a basis of calculation because this proportion has been used in the past.

### 4.3. Relation to Experimental Design

One reason for using at least two sights chosen at random from each lot is in order to obtain a ready measure of variation within the lot, so that excessive differences can be observed with the help of the yardsticks provided for the purpose and attention can thus be directed to possible heterogeneity in the productive process that might lead to deterioration. Another reason is the desirability of averaging together results from different sights of a lot in order to eliminate some of the chance variation and make trends and excessive errors more clearly discernible.

These two reasons are features of the modern theory of the design of experiments, which rests on the twin principles of replication and randomization.* Replication (that is, repetition) has two values. One of these is the increased accuracy obtainable by averaging the results of several observations. The other, which is of the greater importance though perhaps less obvious, is the provision of data for estimating the variance due to errors of observation, data consisting essentially of the differences between the results obtained from experiments in which the controlled conditions are the same. Knowledge concerning this error variance is necessary before any conclusion can be drawn, since real effects can only be distinguished from chance when information is available as to the distribution of the chance factor. It is true that estimates of chance variance can sometimes be drawn from other analogous experiments or from general knowledge; but such estimates are of doubtful validity when applied to a new set of data that must inevitably differ somewhat in origin from those provided by older experiments. The most satisfactory estimate of variance is that provided by the *self-contained* experiment, which is arranged from the beginning so as to

* See FISHER, *Design of Experiments*, especially Secs. 9 and 10 of Chap. 2, Sec. 20 of Chap. 3, and Secs. 25 and 26 of Chap. 4.

yield such an estimate at the same time as the measures of the main effects to be investigated.

Randomization is necessary in order to make the results valid, in the sense that the calculated estimates are free from bias. When the selection of cases for particular kinds of treatment is made by any human being, and especially by a person familiar with the things selected, there is grave danger of bias, conscious or unconscious, in the assignments. Zeal and conscientious care in the selection are no safeguards against bias, as has been demonstrated repeatedly in a variety of fields. The principal reason for this is that attempts to get a "representative" sample or set of conditions by conscious choice usually result in excessive variance within the group in a way that cannot be taken account of by any possible calculation when the probabilities necessary for testing the significance of the results are sought. Such difficulties can be overcome only through the actual utilization of the mechanisms of games of chance or their equivalents. The general framework of the experiment should be established by careful design; but within this framework the allocation of the particular treatments to individual cases must be strictly at random.

When applied to the sampling inspection of articles coming successively from a production line, the principles of replication and randomization suggest the following procedure: If 10 percent, for example, of the product is to be subjected to a specially intensive testing, with a fixed number chosen from each successive batch of some chosen size, these principles call for the selection at random, within each batch, of at least two articles. Thus, samples of 10 per cent taken in accordance with these principles require batches of at least 20 articles.

A seemingly more natural and perhaps a more usual practice when a 10 percent sample is to be taken from a production line is to take every tenth article. If there should be a cycle of period 10 in any part of the productive process, this practice would introduce a biased error into the averages. It also fails to provide a continuing estimation of variance and therefore encourages a reliance on the variance obtained from data available when the system was inaugurated, observations which may in some cases be unfortunately few in number, and which always refer to a past time and are therefore more or less obsolescent.

### 4.4. Efficient Formulas for Utilizing Results

The formulas given in preceding parts of this chapter have properties of maximum efficiency and accuracy of interpretation, provided that the actual distribution of bomb coordinates under homogeneous conditions is of the bivariate normal form. This form of distribution is strongly suggested by general mathematical theory based on the idea of the errors

as resultants of a large number of independent causes. It is approximated by various records of actual bombing, the principal deviation being a slight skewness of the range errors due to a small proportion of defective bombs that tumble and fall short. For distributions having moderate deviation from normality the statistical efficiency of the methods based on the assumption of normality is still nearly 100 percent.

For the normal distribution in one or more variates, it is known that all the information supplied by any random sample with respect to the unknown parameters of the distribution is contained in the sample means and mean squares of the individual variates, and in the mean products of pairs of variates. Only these functions of the observations, and functions of these functions, are capable of full efficiency in estimation of the unknowns. If any other function of the observations is used for estimation or for the testing of any hypothesis regarding the unknown parameters, there is a reduction in accuracy. If the use of an inefficient statistical method is insisted upon, the loss of accuracy can be made good by increasing the number of observations in a proportion inverse to the efficiency of the method in question.

These general principles are illustrated by the use of the mean deviation, or mean of absolute values of errors, as a measure of dispersion, by comparison with the standard deviation, or by the median in comparison with the mean. If a variate $x$, for example the deflection error of bombs dropped under certain conditions, is normally distributed, this means that the element of probability is

$$\frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\xi)^2}{2\sigma^2}}\, dx$$

that is, the integral of this expression between any two limits is the probability of $x$ falling between these limits. This probability distribution involves two parameters, namely, the mean value $\xi$ and the standard deviation $\sigma$, measuring respectively, in the example of deflection errors, the deflection component of the mean point of impact of all bombs dropped under the specified conditions, and dispersion about this mean point of impact. The square of the standard deviation is known as the "variance"; this parameter $\sigma^2$ is the mean value of the squares of deviations from the mean.

When we have $N$ observations $x_1, x_2, \ldots, x_N$ of the variate $x$, we may calculate from them various functions, each known as a "statistic," designed to estimate the unknown parameters $\xi$ and $\sigma$. Thus $\xi$ may be estimated by the arithmetic mean

$$\bar{x} = \frac{x_1 + \cdots + x_N}{N}$$

or alternatively by the median $m$, which is defined as the middle observation (when the number $N$ is odd) when the observations are arranged in order of size. ˙A set of $N$ independent observations is called a "sample." If from many samples of $N$ we form the means $\bar{x}$, these will have a distribution of their own, which happens to be normal also. The mean of this new normal distribution has the same value $\xi$ as the old. But instead of the variance $\sigma^2$, the distribution of means has for its variance the reduced value $\sigma^2/N$. This reduction in variance measures the greater accuracy in averages by comparison with single observations.

The distribution of medians from samples of $N$ is not of the normal form but approximates it if $N$ is large. The variance of this distribution of medians is less than that of $x$ but greater than that of the means $\bar{x}$. The variance of medians is in fact approximately $\pi/2$, or 1.5708, times the variance of means from samples of the same size when $N$ is large. The inefficiency of the median is expressed by this greater dispersion, or by considering the relative sample numbers required to obtain the same accuracy by the median as by the mean. Taking the latter point of view, and letting $N_0$ be the sample size to be used when the arithmetic mean is taken as the estimate of $\xi$, while $N_1$ observations (for example, bombs) are to be used when the median is taken instead of the mean, we have for the two variances

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{N_0} \qquad \sigma_m^2 = \frac{\pi\sigma^2}{2N_1}$$

If the two variances are to be equal, the ratio of the sample numbers is

$$\frac{N_0}{N_1} = \frac{2}{\pi} = 0.6366$$

Consequently, the efficiency of the median is said to be 63.66 percent. The efficiency of the mean has been demonstrated to be 100 percent; that is, its variance has the smallest possible value.

The situation is similar with respect to estimates of variance. There are infinitely many statistics whose mean values are equal to $\sigma^2$. Among all these, the particular statistic

$$s^2 = \frac{S(x - \bar{x})^2}{N - 1}$$

(where $S$ stands for summation over the sample) enjoys the property that its distribution has a smaller variance than any of the others. In particular, when the mean error (that is, the arithmetic mean of the *absolute values* of the deviations from the mean) is multiplied by a factor such that the mean value of the resulting statistic is $\sigma$, this statistic has a distinctly wider dispersion than $s$ when the sample size is the same. Both

these statistics have distributions approaching normality of form when $N$ is large, with the efficiency ratio

$$\frac{1}{\pi - 2} = 0.87597$$

The wide current use of the mean deviation thus has an efficiency of only about 87.6 percent. If bombs are dropped for the purpose of ascertaining their dispersion, the use of the mean deviation is equivalent to wasting about one bomb in eight.

Against this waste of bombs may be set the fact that the mean deviation is easier to calculate than the standard deviation $s$. The labor of calculating $s$ is, however, much less than is sometimes supposed and is trivial if a calculating machine of any kind, or even a table of squares, is at hand. The identity

$$S(x - \bar{x})^2 = Sx^2 - \bar{x}Sx$$

means that the individual deviations from the mean need not be computed. In view of the high cost of aerial bombing, it appears definite that the use of mean deviations, and of functions of them, ought to be eliminated.*

### 4.5. Exact Probabilities

The standard deviation $s$ defined above has another advantage over the mean deviation and all other methods of estimating $\sigma$ in that the exact distribution of $s$ for samples of every size, even the smallest, is a known elementary mathematical function, and tables of its probability integral are available. It also has numerous other practical advantages in more complicated situations, such as those involving least squares or the analysis of variance. In view of the small samples that must sometimes be used, the property of having readily determinable exact probabilities is a valuable one. Asymptotic formulas are somewhat dubious in application to particular cases unless a precise limit of error is at hand and are thus inferior to exact probabilities.

The choice of suitable statistics is also influenced by another and more subtle criterion of accurately determinable probabilities. This is the desideratum that in testing any hypothesis we should use probabilities that depend on nothing but the observations, together with any numbers that may be specified in the hypothesis. Statisticians have often, through lack of existence or knowledge of better methods, used formulas for standard errors and other measures of accuracy in which were involved unknown parameters. The practice has been to substitute

---

\* See FISHER, "Mean Error and Mean Square Error." For a discussion of the general theory of the efficiency of statistics, see Fisher, "Theory of Statistical Estimation," and "Statistical Theory of Estimation."

for these parameters estimates of them obtained from the sample. This substitution however entails an additional sampling error not under control. Modern statistical theory has succeeded in devising various ways of overcoming this difficulty. The statistical methods recommended in this chapter all fulfill the criterion that the associated probabilities are exact, free from unknown nuisance parameters. This is true in particular of the measures $T_O$, $T_M$, and $T_D$ combining range with deflection errors. If the new observations are from the same population as the old and if this has a normal distribution, then the distribution of each of these $T$'s is absolutely independent of all unknown parameters. Only the sample numbers are involved, and these are of course known. To pass from the observations to the relevant exact probabilities through these $T$'s is a purely mathematical operation, involving no element of judgment. It is true that the available tables do not cover fully the $T$ distribution, so that asymptotic approximations must be used for the present, but this is a defect that can be remedied with only a little more mathematical research and calculation. The probabilities respecting $T_M$ can be found easily and accurately from existing published tables when only one flight is involved, but the combination of these values for different flights leads to a more general distribution,* which is also applicable to $T_D$ and $T_O$.

## 4.6. Reasons for the $T$ Method of Combining Range and Deflection Errors

The particular method of combining range with deflection errors represented by the $T$ formulas may be questioned. To many it will seem more natural to use measures, apparently simpler, based on the simple distances from the bombs to the target. After all, it will be said, the ultimate value of a bomb aimed at the enemy depends only on the simple geometrical distance from its point of impact to the target, and not on any figurative measure of distance that is constant along ellipses instead of circles. The simple distance has a clear meaning, intelligible to everyone. It can be measured on a map without reading off coordinates. Why not, then, use a criterion of excellence of bombing depending, in an inverse sense, only on the distances of the bombs from the target?

The several distances might be combined by simple averaging, or by taking the mean square, or in numerous other ways. Use of the mean square, or the sum of the squares, of the distances, has a special attraction in that the sum of the squares of the distances of the bombs from the target equals the sum of the squares of their distances from the mean point of impact, plus the product of the number of bombs by the square of the distance from the mean point of impact to the target. This

---

* This mathematical gap has been filled in since the above text was written. The results are expected to appear in the *Annals of Mathematical Statistics.*

relation, which is analogous to

$$T_O^2 = T_M^2 + T_D^2$$

makes it possible to examine separately two components of the total sum of squares that may be due to quite different causes. Furthermore, estimates of the circular probable error and similar parameters have greater statistical efficiency, that is, smaller standard errors for a given number of observations, if based on the sum of squares instead of the simple sum or mean of the distances.

But the efficiency even of the mean square circular error for estimating the general degree of dispersion is less than that of $T_O$ or $T_D$, provided the base period sample is representative of the true distribution as regards the ratios of the variances of range and deflection, and as regards the correlation. If the true mean point of impact is unknown or if we choose to ignore it, the best available measure of general degree of dispersion is $T_D$ or a function of $T_D$. If the true mean point of impact could be assumed to be at the target, the best measure of general degree of dispersion would be $T_O$ or a function of $T_O$. Such measures have smaller variances than measures of the same parameter based on the mean square circular error in a ratio which, if we can trust the experience described in Sec. 5, is slightly less than 82 percent. This implies that the estimation of dispersion can be carried out with about equal accuracy with 82 bombs if $T_D$ or $T_O$ is used, or with 100 bombs if the mean square circular error is used. The simple mean of distances from the target or from the MPI, or any estimate involving mean deviations, would require distinctly more than 100 bombs to gain the same accuracy obtainable with 82 bombs when $T$ is used.

The seemingly paradoxical result emerges that, to obtain the best possible estimate of the mean square deviation from the target to be expected in later bombing under like conditions, the mean square deviation from the target in the observed cases should not be used. Likewise, to estimate the mean deviation from the target as accurately as possible for the future, the mean deviation from the target in the past should not be used. The same is true of shooting and of all other situations in which the joint distribution of two or more variates does not have circular or spherical symmetry. Indeed, for any given set of ratios among the variances and covariances, the mean square deviation from the target and the mean deviation from the target for the population of bombs or shots, from which our observations constitute a sample, are definite multiples of a parameter measuring general degree of dispersion; and such a parameter is best estimated by a quadratic expression of the $T^2$ type.

For detection of bias in the mean point of impact, the use of $T_M$

has advantages over all other methods that have been studied.* It is true that when only a bias in a specific direction, such as that of range errors, is to be considered and it is certain that there is no bias in any other direction, the mean of the components of deviation (positive and negative) in the particular direction is the most efficient measure, while the use of a measure of bias in all directions weakens somewhat the test for a bias in the particular direction. This difficulty is, however, inescapable when there is any real importance and likelihood of errors of bias in more than one direction. If a rule were established that a bomb should be classed as "bad" if its range error were such as to be exceeded with probability 0.01 and it should also be classified as bad if the deflection error were such as to be exceeded with probability 0.01, there would be some danger of misunderstanding. Many persons would suppose that the probability of a good bomb being wrongly classified as bad under such circumstances is 0.01, whereas actually this probability is in the neighborhood of 0.02. Such erroneous ideas abound in statistical practice in many fields. Use of the combined measure $T_M$ avoids this difficulty.

The $T$ statistics are algebraic invariants and as such possess certain desirable properties. Any transformation of coordinates such as a rotation of axes or a change in units of measure, if applied both to the old and to the new observations, leaves each of the $T$'s invariant and therefore does not affect the probabilities involved. This same property of invariance is also important in obtaining the relevant probability formulas. It implies the very important fact that the probabilities associated with the $T$'s are exact numbers, not functions of nuisance parameters whose values are never exactly known.

There may be a thought that instead of $T$ another invariant known as the "generalized variance" might be used.† The generalized variance has many important potential uses. However, it is not so suitable as $T$ for the purposes here in view because, in two dimensions, it is essentially a measure of distance (in a generalized sense) from a *line* instead of from a point. While $T_M$ and $T_O$ can be zero only if all the bombs fall on the target, and $T_D$ can be zero only if all the bombs fall on one point, the generalized variance will be zero if all the bombs fall on any one line, no matter how they are scattered along it. This consideration also eliminates for the present purpose several other statistics obtained by maximum likelihood methods.‡

An objection likely to be made to the $T$ formulas is against the inclusion of the term in the product of range by deflection errors. If the

* See HSU, "Generalized $T$"; WALD, "Tests Concerning Several Parameters."
† WILKS, "Generalizations in the Analysis of Variance."
‡ PEARSON and WILKS, "$k$ Samples of Two Variables."

correlation between range and deflection in the sample in the base period were exactly zero, this term would be zero.   It may be argued on a priori grounds that the true correlation is zero, and that when, as in the sample here used, the observed correlation is only slightly greater than its standard error (Sec. 5 gives, after adjustments, a correlation of $-0.07535$, which is 1.2 times its standard error), later work should proceed on the fixed assumption that the correlation is zero.   A further extension of this idea might lead to ignoring the difference between the variances of range and deflection, and consequently using the mean square deviation, whose inefficiency has been described above.

While omission of the small product term from $T^2$ would seem at first to save a little computational labor, it would lead to great mathematical complexities in connection with the probabilities.   The resulting function is not an invariant, and its distribution involves as a nuisance parameter the ratio of range variance to deflection variance in the population, which is estimated by the sample only with an uncertainty that makes highly dubious any probabilities calculated on this basis.   Moreover, if it should turn out that there really is some correlation between range and deflection errors, whether the causes and magnitude of such correlation are known or unknown, this correlation would constitute an additional nuisance parameter having serious effects on the accuracy of judgments based on the observations.   In this connection it should be recalled that, while the particular sample here considered does not establish existence of a correlation, other reports seem, at least superficially, to show high and significant correlations between range and deflection errors in practice bombing.   Some officers are of the opinion that a real correlation exists.

Another way of avoiding the calculation of products of range by deflection errors would be to make separate tests of significance for range and deflection, using the variance ratio distribution, and then to combine these by the method of combining independent test probabilities propounded by Fisher and studied by Wallis and others.*   This method requires use of the exact probabilities rather than of the values of the variance ratio at a few points such as those for probabilities 0.01 and 0.05.   Consequently, the tables of the $F$ distribution must be used twice for every such test and thus for every point plotted on control charts corresponding to those for $T_M$, $T_D$, and $T_O$.   The labor involved would amount to far more, and the assumptions necessary for validity are more stringent, by this than by the $T$ methods.

The methods based on $T$ are valid and accurate whether a true correlation exists or not, and whatever the ratio of the true variances may be.

---

* FISHER, *Statistical Methods*, Sec. 21.1 of Chap. 4; WALLIS, "Compounding Probabilities."

The only limitations on this statement are in the assumptions of normality and independence, which are made throughout, not only in the present discussion but in virtually the whole literature of bombing and firing, including that concerned with methods of combining observations. These assumptions, in relation to test bombing, are discussed in Sec. 3.1.4, pp. 146–147.

## 5. STATISTICAL ANALYSIS OF CERTAIN TEST BOMBINGS

### 5.1. Summary of the Data

In order to obtain variances in range and deflection, and a correlation, from which the coefficients $a$, $b$, $c$ in $T^2$ could be found, a series of 65 flights was analyzed. These consisted of all those flights at a certain proving ground between November 4, 1942, and December 23, 1942, during which exactly four bombs were dropped. The coordinates of the point of impact of each bomb with respect to axes through the target, of which one is an average track for the whole flight, were supplied by the proving ground, together with the heading of the airplane when the bomb was dropped and other data. Two of the bombs on each flight were dropped while the airplane was moving in a generally northerly direction and the other two while it was flying in a southerly direction. The actual headings, however, varied considerably, the difference between the two in the same general direction on the same flight being commonly several degrees and in one case amounting to 25 degrees. We therefore applied to the coordinates received a rotation designed to give the errors in range and deflection with respect to the actual heading of the airplane. Table 3.1 gives the principal statistics summarizing both the original data on range and deflection and the values as adjusted by these rotations. Except for the first column of this one table, all numerical results given hereinafter will refer to the rotated values. In each case $r$ denotes the error in range and $d$ the error in deflection. Positive values of $r$ refer to bombs overshooting the target, negative values of $r$ to bombs that fall short. Deflection errors to the right are considered positive; those to the left are negative. The unit of measure is in each case the mil, about $1/1,000$ of a radian, or about $2/35$ of a degree. The data as received refer to actual impacts, uncorrected for differential ballistic wind, nonstandard altitude, nonstandard air speed, or range component of crosstrail. The actual altitudes are understood to have fluctuated rather widely about 7,000 ft., but since the errors are measured in mils the effects of these altitude variations on the results of the analysis should not be great.

If any one rotation had been applied to all the points of impact, the squared distance $\bar{r}^2 + \bar{d}^2$ of the mean point of impact from the target would be invariant. The fact that $\bar{r}^2 + \bar{d}^2$ has two different values, 8.83

for the original and 9.67 for the rotated data, results from the differences of the rotations applied for the different bombs.

TABLE 3.1
SUMMARY OF BOMBING ERRORS IN 65 FLIGHTS (260 BOMBS)

|  | Original (mils) | Rotated (mils) |
|---|---|---|
| Mean range error* (over), $\bar{r}$................... | 0.5108 | 1.3450 |
| Mean deflection error* (right), $\bar{d}$. ......... | 2.9292 | 2.8046 |
| Standard deviation of range errors, $s_r$ ........ | 6.8596 | 6.9074 |
| Standard deviation of deflection errors, $s_d$..... | 8.7195 | 8.6191 |
| Correlation of range with deflection errors..... | --0.037 | −0.035 |

* These are means of positive and negative errors, not of their absolute values.    The first two lines in Table 3.1 would give the coordinates of the mean point of impact (MPI) of all the 260 bombs if the same course had been flown in each case.  They reflect biased error, whereas the next two lines give measures of the unbiased or compensating errors, often measured less accurately by the "mean deviation," or mean of absolute values of errors, which to be comparable with the standard deviation, would need to be multiplied by a factor that for large samples approximates $\sqrt{\pi/2} = 1.253$.

Both the mean range error and the mean deflection error (that is, the means of the actual values of the errors, positive and negative, not of the absolute values) calculated from the rotated data are statistically significant relative to their standard errors.    Thus it is clear that there were biased errors throwing the mean point of impact (MPI) beyond and to the right of the target.    The bias in deflection could have been detected from the original, unrotated observations, but in them the bias in range, being smaller, was obscured by the addition of the variance among the four headings on the same flight to the variance due to other causes.

As in various other bombing records, the standard deviation in deflection is greater than that in range.    The excess is 24.8 percent in the present sample and is statistically significant.    The appropriate test is in terms of $F = (s_d/s_r)^2$, which equals 1.6158 for the original and 1.5570 for the rotated data.    With the 259 degrees of freedom in each of the variance estimates, the probability that $F$ should exceed 1.470 is 0.001.    When, as here, the greater variance is divided by the smaller, the probability is less than 1 in 500 of obtaining so great a ratio of standard deviations as that actually observed if the true standard deviations are equal.    At the customary 0.05 probability level, confidence limits for the percentage by which the standard deviation in deflection exceeds that in range are 12.6 percent and 38.2 percent.

The correlations −0.037 and −0.035 are completely insignificant, since their standard error is $1/\sqrt{259} = 0.0621$.    This of course does not definitely prove that range and deflection errors are mutually independent; it only fails to disprove it.

Among the 65 flights there were variations with respect to both personnel and equipment. Ten different teams, each consisting of a pilot and a bombardier, were represented. The number of flights made by a team varied from 1 to 17 in the several cases. Automatic Erection System (AES) was used in 24 cases and Stabilized Bombing Approach Equipment (SBAE) in 8 cases. The distribution of the flights in relation to personnel and equipment is shown in Table 3.2. The first columns designate the pilot and the bombardier by their initials and indicate the combination of the two by a number.

TABLE 3.2
DISTRIBUTION OF FLIGHTS BY PERSONNEL AND EQUIPMENT

| Pilot | Bombard-ier | Team number | Number of cases | | | | |
|-------|-------------|-------------|-----------------|-----|------|-----|-----|
| | | | Neither AES nor SBAE | AES only | SBAE only | Both AES and SBAE | Total for team |
| WT | H | 1 | 10 | 5 | 2 | | 17 |
| W | S | 2 | 6 | | 1 | 3 | 10 |
| R | R | 3 | 4 | 3 | | | 7 |
| W | H | 4 | 1 | | 1 | 1 | 3 |
| R | H | 5 | 7 | 4 | | | 11 |
| K | U | 6 | 1 | | | | 1 |
| K | H | 7 | 4 | 4 | | | 8 |
| K | L | 8 | 2 | 2 | | | 4 |
| W | R | 9 | 1 | 2 | | | 3 |
| K | R | 10 | 1 | | | | 1 |
| Totals................... | | | 37 | 20 | 4 | 4 | 65 |

## 5.2. Analysis of the Data

The primary object of the analysis was to obtain suitable coefficients $a$, $b$, and $c$ for use in the expressions for $T^2$ that are to measure the accuracy of individual bombsights. However, the variances entering into these expressions are presumably affected by the variations of equipment and of personnel among the flights examined. It is to be expected that this heterogeneity will increase the sample variances above the values they would have if equipment and personnel had been kept constant on all the flights. Such an increase of variance would reduce the values of $a$, $b$, and $c$, and therefore the values of $T^2$ computed from them in future tests, giving an illusory impression of accuracy in those tests in which equipment and personnel are close to the average situation in the cases analyzed. To determine whether this is a serious bias and to correct for it if it is, it seemed desirable to consider regression equations for predicting the range and deflection errors in terms of equipment and

personnel.   At the same time, the opportunity was at hand to obtain
tentative estimates characterizing the performance of particular teams
and of the particular types of equipment and to test the accuracy of
these estimates.

Several different regression equations were used.   In the first instance,
to make a rather extensive elimination of extraneous factors, we con-
sidered regression formulas with no less than 79 terms.   Besides an
additive constant, these terms were

$x_1$, defined as 1 if AES was used, and otherwise 0

$x_2$, defined as 1 if SBAE was used, and otherwise 0

$x_3$, defined as 1 if the airplane was flying north, but 0 if it was flying
south

$x'_i$, ($i = 1, 2, \cdots , 10$), equal to 1 if the bomb was dropped by the
$i$th team, and otherwise 0.   The number of variables in this
group was of course 10, the number of teams.

$x''_{ij}$, ($i = 1, 2, \cdots , 10$), equal to 1 if the bomb was dropped on the
$j$th of the flights of the $i$th team, and otherwise 0.   The number
of variables in this group was 65, the same as the number of flights.

Two different predictands, the range and deflection errors, were used with
this same set of predictors.   The regression equation for estimating
range error in terms of all the predictors was of the form

$$R = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + \sum_{i=1}^{10} b'_i x'_i + \sum b''_{ij} x''_{ij}$$

with a similar formula for estimated deflection error $D$.   The coefficients
were determined by the method of least squares.

It might seem that the solution of a least-square problem with 79
unknowns would be an utterly fantastic task, to say nothing of two such
solutions.   Yet through special circumstances it was possible not only to
carry through this calculation completely, with the relevant standard
errors and tests for significance in terms of probability, but to do so
with a very modest amount of computing.   One of these circumstances
was that, because of the nature of the variables, a large majority of the
elements of the matrix of the normal equations were zero.   Also, a
simplification was obtained through choice of 13 linear relations that
can be imposed somewhat arbitrarily upon the regression coefficients
because of 13 identical linear relations among the predictors, which
reduce the rank of the 79-rowed matrix to 66.   These 13 identical rela-
tions express the following facts:

The sum of the values of $x''_{ij}$ for all the flights of the $i$th team equals
$x'_i$.   There are 10 of these relations, corresponding to the 10 teams.

The sum $x_1' + x_2' + \cdots + x_{10}' = 1$
The sum of the values of $x_{ij}''$ for all flights using AES is $x_1$
The sum of the values of $x_{ij}''$ for all flights using SBAE is $x_2$

A linear constraint corresponding to each of these was placed upon the regression coefficients in each group in order to make them determinate. Though it is customary in problems of this kind to use the simplest constraints by merely equating to zero the sum of the coefficients in each group, we chose a different set of constraints adapted to simplifying the numerical work. For example, where a constraint of the form

$$b_1' + b_2' + \cdots + b_{10}' = 0$$

would ordinarily be imposed, we used instead

$$17b_1' + 10b_2' + 7b_3' + \cdots + b_{10}' = 0$$

with coefficients equal to the numbers of flights of the respective teams. This made possible a great reduction in labor because of the occurrence of this particular linear combination in the normal equations. All results of importance involving the $b''$s will necessarily be in terms of their differences and do not seem to be affected by replacing one such constraint by another.

The solution of the normal and constraining equations was carried out literally, in terms of symbols for the right-hand members of the normal equations, instead of numerically. In this way it was easy to obtain the two regression equations, one for range and the other for deflection errors, by substituting the two sets of right-hand members. Furthermore, the literal solution made easy the determination of standard errors and correlations among the regression coefficients in either equation. Thus it played somewhat the same role as that of the inverse matrix in least-square problems involving no identical relations or constraints.

The question whether there was significant variation among the 10 teams with regard to bias in range and deflection errors was tested by means of the variance ratio distribution. The denominator of this ratio was the estimate of variance based on the sum of squares of residuals from the regression equation. Since there were 260 observations and 79 regression coefficients connected by 13 constraints, the number of degrees of freedom for this residual variance was $260 - (79 - 13) = 194$. The numerator was a quadratic form in the $b''$s with 9 degrees of freedom, since they are subject to the one relation indicated above. The value of such a quadratic form can be found by first ignoring any one of the $b''$s (which we arbitrarily chose as $b_1'$), inverting* the covariance matrix of

* Several methods of numerical inversion of matrices are described in Dwyer, "Solution of Simultaneous Equations" and Hotelling, "Matrix Calculation." The method given by Dwyer is particularly valuable.

the other nine corresponding to unit variance of residuals, and then calculating straightforwardly the quadratic form of which this inverse is the matrix and in which the variables are the nine $b''$s used. This has only to be divided by 9 to give the numerator of the variance ratio, essentially an estimate of variance among the $b''$s, but differing from a mere sum of squares of their deviations from their mean because of their unequal weights and correlations resulting from the unbalanced distribution of equipment and personnel. The inverse matrix is the same for range as for deflection errors, since it depends only on the distribution of the predictors and not on the predictands.

For range errors the variance ratio $F$ found in this way was 1.423; for deflection errors it was 1.426. These are considerably less than the value 1.923 of $F$ that would just be regarded as significant on the customary 0.05 level and indeed are little greater than the value that would be exceeded by chance with probability 0.20 if the true mean range and deflection deviations were the same for all teams. Thus no evidence pointing to any variation among teams in these biased errors emerges from this test.

With two variates such as range and deflection errors the two separate tests may, with a substantial increase in accuracy, be replaced by a single test. In place of the foregoing variance ratios we may well consider a combined test due to Wilks.* Each variance is replaced in this test by a symmetric two-rowed determinant. The elements of the denominator determinant are the variances and the covariance of range and deflection as estimated from the residuals from the two regression equations in which they are the predictands. The numerator determinant's elements are the two quadratic forms computed from the two sets of $b''$s, and the bilinear form with the same matrix and involving both sets of $b''$s. The value 1.696 obtained for this generalized variance ratio would be exceeded by chance with a probability considerably greater than 0.05, though much further below 0.20 than in the separate variance ratios for range and deflection. Thus even this more sensitive test fails to give real evidence of variation in mean point of impact among the men engaged in dropping the bombs. That there is a real bias, at least in deflection, is easily inferred from Table 3.1 or from other bombing records. Apparently the bias is much the same for all the 10 teams on whose performance these records are based.

Application of the corresponding tests for the effect on the MPI of AES, of SBAE, and of direction of flight was much simpler because each of these effects corresponds to only 1 degree of freedom, so that the test for either range or deflection reduces merely to the comparison of a regression coefficient with its standard error. When AES was used, the MPI

---

* WILKS, "Generalizations in the Analysis of Variance."

was 1.6536 mils farther in the "short" direction than as if it had not been used, according to the regression coefficient, which makes allowances for effects of SBAE, direction of flight, and personnel. This coefficient is statistically significant on the customary 0.05 probability level. Since the MPI for all the bombs was 1.3450 mils "over," the use of AES appears to have been advantageous in this respect. The corresponding coefficient for deflection was not significant, and the combined test by Wilks's ratio of determinants also failed to yield a significant result on the 0.05 level. The same tests applied to SBAE gave results far from significant, suggesting no difference according as this equipment was used or not.

The effect of direction of flight on MPI was tested by means of the coefficient $b_3$ in the 79-term regression equation. For range the effect was insignificant. But for deflection the value $-4.1045$ was significant even on the 0.01 level. This indicates that the bias toward the right was considerably less flying north than flying south.

The variations of MPI among flights were tested by an approximate method, simpler than that used for personnel variance, which would for 65 flights with 53 degrees of freedom have involved the considerable labor of inverting a matrix of 53 rows. Such a piece of calculation is not so completely out of the question as would once have been thought, but it did not seem worth while for the present purpose. By ignoring the unbalance occasioned by disproportionate frequencies of equipment and personnel, while utilizing the balance of north and south flights and the fact that exactly four bombs were dropped on each flight, these considerations lead to the estimate $4\Sigma b''^2/53$ of variance among flights. This turns out to be insignificant for deflection but significant for range. The combined test for range and deflection failed to yield a significant result. Thus any important effect on MPI of variations among flights seems to be confined to the range components. Such variations refer to effects of varying weather and other factors not taken into account in the regression equation but do not include the variations of personnel and equipment considered.

Since it is to be expected that pilots and bombardiers will change from time to time, as will also the weather, the variances of deviations from the 79-term regression functions will be less relevant than if the effects of the fluctuations had not been eliminated. If the personnel were to remain constant, and of constant ability, and if an adequate measure of seasonal variation were available from records over a period of years based on ore-filled bombs, the appropriate estimates of variances and correlation for appraising bombsights under test would be in terms of deviations from a regression equation involving all these factors. The omission of the terms in the regression function corresponding to per-

sonnel may also be justified on the ground that, as seen above, no discernible differences between the MPI's of different teams have been found.

Inclusion of an allowance for equipment type does seem to be needed, first because of the highly significant difference in MPI associated with AES, and second because of the prospect that future standardization will lead either to the use or the nonuse of each of the types of equipment with all sights tested alike. Furthermore, it is to be hoped that methods will be found to eliminate the difference between the MPI of bombs dropped on northward and on southward flights. The values recommended for the coefficients $a$, $b$, and $c$ of the test functions $T^2$ to be used for future control charts have therefore been based on the residuals from regression functions with only four terms. Besides a constant, these terms involved

$x_1$, which equals 1 if **AES** was used and is otherwise 0

$x_2$, which equals 1 if **SBAE** was used and is otherwise 0

$x_3$, which equals 1 if heading was north, but 0 if heading was south

These regression equations were found to be, for range errors in mils over,

$$R = 2.9125 - 1.4457x_1 - 2.8149x_2 - 1.3746x_3$$

and for deflection, measured in mils right,

$$D = 5.4099 - 0.8106x_1 - 1.0176x_2 - 4.3615x_3$$

The deviations from either of these regression equations are 260 in number but have only 256 degrees of freedom because of the four constants in the equation, in accordance with well-known least-square theory. The sum of squares of these deviations from one of the equations, divided by 256, gives the estimated variance in range or deflection. In the same way, the sum of products of deviations from the two equations, divided by 256, gives the estimated covariance. These values are

Range variance.................. $s_r^2 = s_{11} = 46.294995$

Deflection variance.............. $s_d^2 = s_{22} = 70.033392$

Covariance..................... $s_{rd} = s_{12} = -4.290599$

From these values a simple calculation gives

$$s_r = 6.8040 \text{ mils} \qquad s_d = 8.3686 \text{ mils} \qquad r = -.07535$$

the estimated values mentioned in previous parts of this chapter of the true standard deviations $\sigma_r$ and $\sigma_d$ and correlation $\rho$. The number 256 of degrees of freedom is large enough to make the estimates $s_r$ and $s_d$ quite reliable. The standard error of each of these standard deviations is only about 4.42 percent of its value. That of $r$ is exactly $\frac{1}{16}$, making

the correlation only 1.206 times its standard error. The probability of so great a value of $|r|$ in the absence of any real correlation is about 0.23, so that the hypothesis of independence between range and deflection is not refuted by these data, though there is a faint suggestion that bombs that fall short tend to fall to the right, while those going over the target tend to fall to the left.

The inverse of the matrix

$$\begin{bmatrix} s_{11} & s_{12} \\ s_{12} & s_{22} \end{bmatrix}$$

is

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

and from the above values we find

$$a = 0.021724 \qquad b = 0.001331 \qquad c = 0.014360$$

the values used in the plan presented in Sec. 2.

## 5.3. Ratings of Equipment and Personnel by Test Bombing

A series of test bombings can be utilized to obtain information not only regarding the sights used, but at the same time regarding other equipment and the men who do the flying and drop the bombs. For each flight $T_O^2$ supplies a measure of the general success, and this is easily analyzed into the components $T_M^2$ and $T_D^2$, corresponding, respectively, to deviations of the MPI from the target and of the bombs from the MPI. By combining the values of $T_O^2$ (or of $T_M^2$ or of $T_D^2$) for all the flights on which a particular kind of equipment was used, or on which a particular man or team flew, and contrasting with the flights not involving the same men and equipment, it is possible to obtain ratings of personnel and equipment. Approximate probability distributions are available for testing the reliability of such ratings, since each sum of independent values of $T^2$ may with little error be regarded as having the $\chi^2$ distribution, and the ratio of two such sums the $F$ distribution. Since several kinds of equipment in addition to the personnel can thus be rated on the basis of the same bombing records, such multipurpose test bombing is relatively very economical and productive of information.

Complete validity of such ratings is, however, essentially dependent on a condition that must be provided *before* the bombings, not afterward. This condition is that, subject to whatever general restrictions are appropriate, the assignment of personnel and equipment to particular flights must be strictly at random. There is no way of doing this except by the use of the actual mechanism of games of chance, or of their equivalent in the form of tables of random numbers. In the absence of

such objective chance, the assignments must be made by human beings. Experience has clearly shown that human beings unaided by gambling equipment are unable to assign different test materials to different experimental trials in such a way as to avoid bias, either in the averages to which the experiment leads or in the measures of error that must be computed as functions of deviations from the observed averages. Consequently, the ratings of the pilots and bombardiers in Table 3.4, while historically accurate, are subject to the possibility of being explained away by features of the method by which assignments were made. This same difficulty qualifies the statistical interpretation of economic time series and other such records which, though frequently the best available guides, lack the kind of validity possible only in a well-designed experiment.

To secure *efficiency* as well as *validity* in an experiment, it is usually necessary to impose restrictions on the randomization. Thus in comparing the results of two treatments when the chance variances are the same in both cases, the standard error of the difference of the means is proportional to $\sqrt{1/N + 1/N'}$, where $N$ and $N'$ are the numbers to which the two treatments are applied. If $N + N'$ has any fixed value, the standard error is therefore a minimum when $N = N'$. Randomization in assignment of the cases to the two treatments should then be subject to the condition $N = N'$. This could be done by shuffling and randomly dividing a pack of cards consisting of black and red cards in equal numbers. It could not be done merely by flipping a coin once for each case, since the condition $N = N'$ might not then be satisfied. This question of efficient arrangement of experiments through restricted randomization has many ramifications, particularly where several questions, such as the efficacy of several kinds of equipment serving different purposes, are to be tested simultaneously in the same experiment.*

Values of $T_O^2$, $T_M^2$, and $T_D^2$ were computed for each of the 65 flights. Graphs of these three functions in the order in which the flights were made are appended as Figs. 3.8, 3.9, and 3.10, respectively. These figures have the general appearance of control charts. They differ, however, from suitable control charts in certain respects. Measures of quality should be plotted on a control chart in the order in which the articles are manufactured, whereas here the order is that in which the tests were performed. In this material several different series of serial numbers were represented, referring to different modifications of design and perhaps different manufacturers. It is best to run a separate control chart for each modification and for each factory.

The points in these graphs are on the average higher relative to the theoretical probability 0.20, 0.05 and 0.01 levels shown than would be

* Fisher, *Design of Experiments*, is the best treatise on this subject.
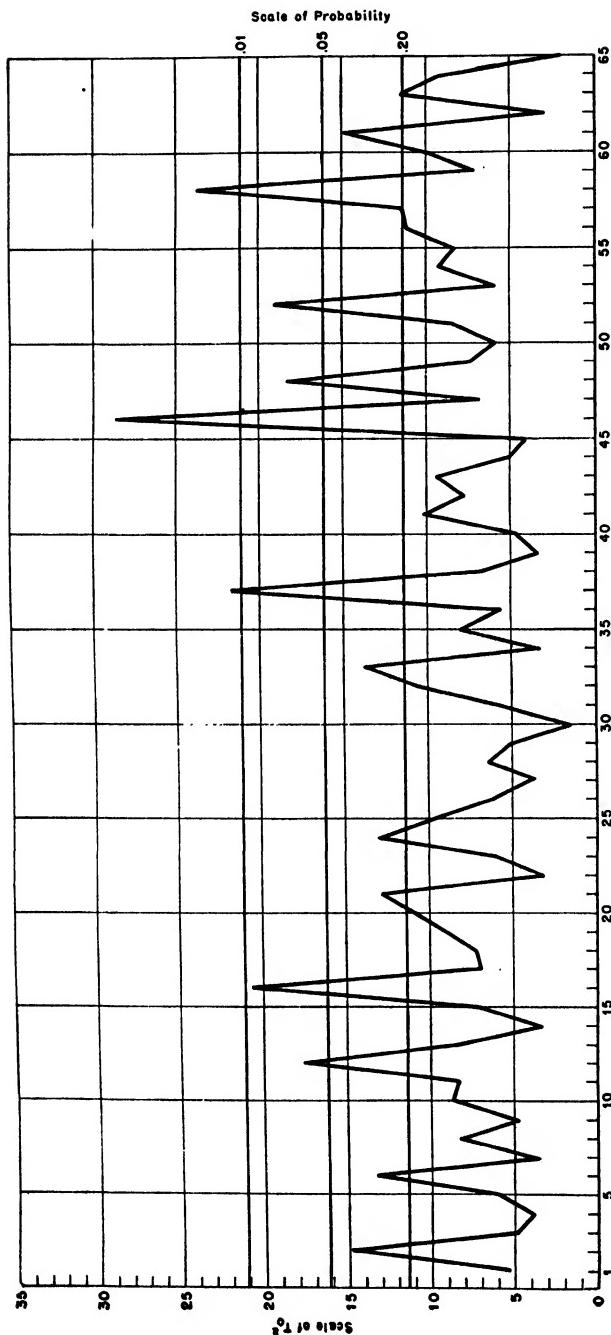
FIG. 3.8.—Over-all measure of error $T_0^2$ applied to sample data.
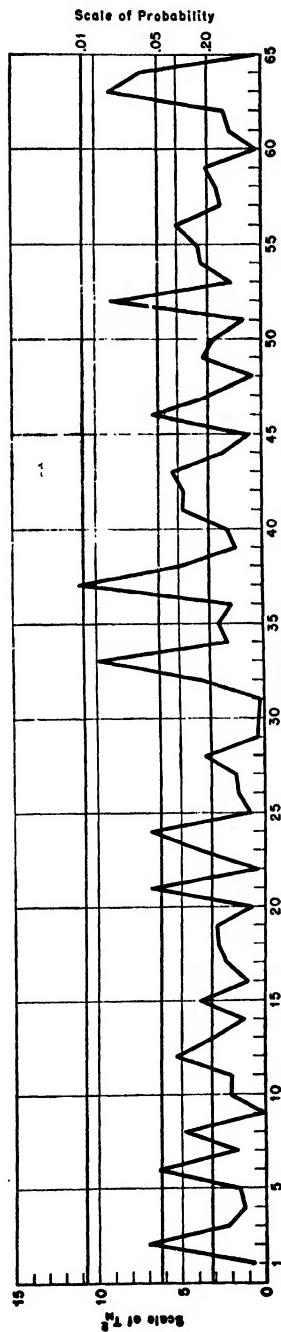
Fig. 3.9.—Errors in MPI in sample, as measured by $T^1_M$.
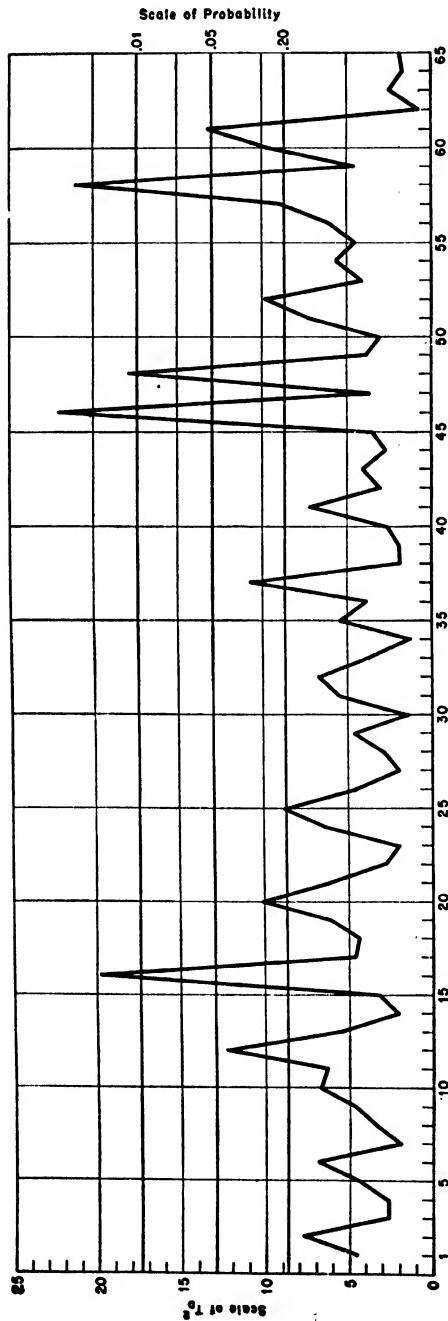
FIG. 3.10.—Dispersion measure $T_D^2$ applied to sample data.

expected in a manufacturing and testing routine in a state of control. The reason for this lies in a particular feature of the way the charts were constructed rather than in any shortcomings of the sights tested. The variances on which the coefficients in $T^2$ are based were chosen in Sec. 5.2 on the basis of deviations from 4-term regression functions, in order to eliminate from these estimates of variance the effects, which it is to be anticipated are not to continue long, of the differences between use and nonuse of the Automatic Erection System and Stabilized Bombing Approach Equipment, and between northward and southward flights. This elimination has the effect of decreasing the variances somewhat and thus of increasing the coefficients in each of the expressions for $T^2$. If $a$, $b$, and $c$ had been calculated by inverting the two-rowed matrix of sums of products of the raw deviations from the target, and if these raw deviations had then been substituted in $T_o^2$, $T_M^2$, or $T_b^2$, the average of the resulting values would for each of these three functions be exactly the number of degrees of freedom, respectively, 8, 2, and 6. If, on the other hand, with the values of $a$, $b$, and $c$ actually used, $r$ and $d$ had been replaced by the deviations from the 4-term regression functions, the averages would again in each case be exactly equal to the number of degrees of freedom. But because actual deviations were substituted for $r$ and $d$ while corrected variances were used in determining the coefficients, the averages exceed the numbers of degrees of freedom. Thus the average of the 65 values of $T_o^2$ is 9.00786 instead of 8. This is reflected graphically in the fact that the proportions of the 65 cases above the horizontal lines are in general somewhat greater than the respective probabilities.

From Table 3.3 it appears that AES and SBAE, instead of improving bombing accuracy, seemed to make it worse. The significance of this apparent effect may be tested (with an approximation equivalent to identifying the sample variances and covariance based on 256 degrees

TABLE 3.3

EFFECT OF EQUIPMENT ON OVER-ALL ACCURACY OF BOMBING

(Excellent bombing is indicated by low values of $T_o^2$)

|  | Number of flights | $\Sigma T_o^2$ | Average $T_o^2$ |
|---|---|---|---|
| Flights using AES............................ | 24 | 228.350001 | 9.51458 |
| Flights not using AES........................ | 41 | 357.161062 | 8.71125 |
| Flights using SBAE.......................... | 8 | 79.554375 | 9.94430 |
| Flights not using SBAE...................... | 57 | 505.956688 | 8.87643 |

of freedom with the true values of these parameters in an infinite population) by means of the variance ratio distribution with numbers of degrees

of freedom equal to eight times (since each flight has 8 degrees of freedom) the numbers of flights with and without the equipment in question. For AES the ratio is

$$F = \frac{9.51458}{8.71125} = 1.09222$$

with $n_1 = 192$ and $n_2 = 328$. This is not at all significant since the probability of a greater value of $F$ is considerably more than 0.20. Actually, the relevant probability is more than double this value because of our selection of the greater mean value of $T_o^2$ to put in the numerator. For SBAE we have

$$F = \frac{9.94430}{8.87643} = 1.12030$$

which, with $n_1 = 64$ and $n_2 = 456$, is similarly insignificant. Thus these results do not prove that AES and SBAE have deleterious effects. They would not actually prove it in a really satisfactory manner even if the results had turned out to be significant, since the assignment of the equipment to particular flights was presumably not on the basis of objective randomization. It is entirely possible that the decision to use or not to use either of these kinds of equipment on each occasion had some relation, conscious or unconscious, to the weather, the altitude, the personnel, or some other variable affecting the accuracy.

The same qualifications apply to the average values of $T_o^2$ for the 10 teams shown in Table 3.4. Such scores are reliable indicators of differences in personal ability only if one of two conditions is satisfied.

TABLE 3.4
VARIATION OF OVER-ALL ACCURACY WITH PERSONNEL
(Excellent bombing is indicated by low values of $T_o^2$)

| Teams | Number of flights | $\Sigma T_o^2$ | Average $T_o^2$ |
|---|---|---|---|
| 1—WT–H | 17 | 184.834959 | 10.87264 |
| 2—W–S | 10 | 73.105782 | 7.31058 |
| 3—R–R | 7 | 62.168030 | 8.88115 |
| 4—W–H | 3 | 30.389824 | 10.12994 |
| 5—R–H | 11 | 72.692086 | 6.60837 |
| 6—K–J | 1 | 3.240875 | 3.24088 |
| 7—K–H | 8 | 98.558949 | 12.31987 |
| 8—K–L | 4 | 22.423547 | 5.60589 |
| 9—W–R | 3 | 28.637496 | 9.54583 |
| 10—K–R | 1 | 9.459515 | 9.45952 |
| All flights........ | 65 | 585.511063 | 9.00786 |

Either the assignment of the men must be suitably randomized, or an exhaustive knowledge of the other causes of variation must be at hand

and must show that these other causes exactly balanced as between the different teams.   Since it is practically impossible to fulfill the second condition, objective randomization in advance is the only practical way of validating scores.

Even under randomization, differences in scores prove nothing unless they are too great to be ascribed to chance.   As between any two teams selected without reference to their scores in this bombing series, the effects of chance may be examined by means of the variance ratio, equal to the ratio of the mean values of $T_n^2$, with numbers of degrees of freedom equal to the numbers of flights of the two teams multiplied by the number of degrees of freedom, here 8, for each flight.   However, it would be an error to compare in this way two teams selected because their scores are the greatest and the least or because they occupy adjacent places when arranged in the order of the scores.   Whether the 10 scores as a whole differ significantly from one another might be determined by the application of such a criterion as Bartlett's.*   The tests might also be refined, if it should seem worth while, by a more complete segregation of the effects of personnel from those of each kind of equipment, here represented by overlapping degrees of freedom, and of interactions. We are dealing here with generalizations of the ordinary analysis of variance, generalizations whose fundamental theory has been developed only part of the way necessary for an efficient analysis of the ramifications of bombing accuracy.

### 6. REFERENCES

BARTLETT, M. S.: "Properties of Sufficiency and Statistical Tests," *Proceedings of the Royal Society of London*, Vol. 160A (1937), pp. 268–282.

CRUM, W. L., and SCHUMPETER, JOSEPH A.: *Rudimentary Mathematics for Economists and Statisticians*, McGraw-Hill Book Company, Inc., New York and London, 1946, 183 + ix pp.

DWYER, P. S.: "The Solution of Simultaneous Equations," *Psychometrika*, Vol. 6 (1941), pp. 101–129.

FISHER, R. A.: "A Mathematical Examination of the Methods of Determining the Accuracy of an Observation by the Mean Error and by the Mean Square Error," *Monthly Notices of the Royal Astronomical Society*, Vol. 80 (1920), pp. 758–770.

———: "The Theory of Statistical Estimation," *Proceedings of the Cambridge Philosophical Society*, Vol. 22 (1925), pp. 700–725.

———: *Statistical Methods for Research Workers*, 4th and later ed., Oliver and Boyd, Ltd., Edinburgh and London, 1932–1943.

———: *The Design of Experiments*, Oliver and Boyd, Ltd., Edinburgh and London, 1935, 252 + xi pp.

———: *Statistical Theory of Estimation*, University of Calcutta, Calcutta, 1938.

———, and YATES, FRANK: *Statistical Tables for Biological, Agricultural and Medical Research*, 2d ed., Oliver and Boyd, Ltd., London and Edinburgh, 1943, 98 + viii pp.

* BARTLETT, "Sufficiency and Statistical Tests."   For an account of Bartlett's test, see Chap. 15, Sec. 2.

GOLDBERG, HENRY, and LEVINE, HARRIET, "Approximate Formulas for the Percentage Points and Normalization of $t$ and $\chi^2$," *Annals of Mathematical Statistics*, Vol. 17 (1946), pp. 216–225.

HOTELLING, HAROLD, "The Generalization of Student's Ratio," *Annals of Mathematical Statistics*, Vol. 2 (1931), pp. 360–378.

———: "Some New Methods in Matrix Calculation," *Annals of Mathematical Statistics*, Vol. 14 (1943), pp. 1–34.

———, and FRANKEL, LESTER: "The Transformation of Statistics to Simplify Their Distribution," *Annals of Mathematical Statistics*, Vol. 9 (1938), pp. 87–96.

HSU, P. L.: "Notes on Hotelling's Generalized $T$," *Annals of Mathematical Statistics*, Vol. 9 (1938), pp. 231–243.

KENDALL, M. G., and SMITH, B. BABINGTON: *Random Sampling Numbers*, Cambridge University Press, London, 1939, 60 + $x$ pp.

MERRINGTON, MAXINE: "Table of Percentage Points of the $t$ Distribution," *Biometrika*, Vol. 32 (1942), p. 300.

———, and THOMPSON, CATHERINE M.: "Tables of Percentage Points of the Inverted Beta ($F$) Distribution," *Biometrika*, Vol. 33 (1943), pp. 73–88.

PEARSON, E. S., and WILKS, S. S.: "Methods of Statistical Analysis Appropriate for $k$ Samples of Two Variables," *Biometrika*, Vol. 25 (1933), pp. 353–378.

PEISER, A. M.: "Formulas for Significance Levels of Certain Distributions," *Annals of Mathematical Statistics*, Vol. 14 (1943), pp. 56–62.

ROY, S. N.: "$p$-Statistics, or Some Generalizations in Analysis of Variance Appropriate to Multivariate Problems," *Sankhya: The Indian Journal of Statistics*, Vol. 4 (1939), pp. 381–396.

SEALY, E. H.: *Rational Sentencing of Proof Results*, (British) Ministry of Supply, Advisory Service on Quality Control, Technical Report No. Q.C./S/6, October 20, 1943.

———: *Acceptance Sampling*. address given to the Industrial Applications Group of the Manchester Statistical Society on May 20, 1944, mimeographed, British Central Scientific Office, July 14, 1944.

Statistical Research Group, Columbia University: *Sequential Analysis of Statistical Data: Applications*, rev. ed., Columbia University Press, New York, 1945.

THOMPSON, CATHERINE M.: "Table of Percentage Points of the $\chi^2$ Distribution," *Biometrika*, Vol. 32 (1941), pp. 187–191.

TIPPETT, L. H. C.: *Random Sampling Numbers*, Cambridge University Press, London, 1927.

WALD, A.: "Tests of Statistical Hypotheses concerning Several Parameters When the Number of Observations Is Large," *Transactions of the American Mathematical Society*, Vol. 54 (1943), pp. 426–482.

———: "Sequential Tests of Statistical Hypotheses," *Annals of Mathematical Statistics*, Vol. 16 (1945), pp. 117–186. Reprints available from Institute of Mathematical Statistics, 116 Rackham Hall, Ann Arbor, Mich.

WALLIS, W. ALLEN: "Compounding Probabilities from Independent Significance Tests," *Econometrica*, Vol. 10 (1942), pp. 229–248.

WILKS, S. S.: "Certain Generalizations in the Analysis of Variance," *Biometrika*, Vol. 24 (1932), pp. 471–494.

# CHAPTER 4

# EFFECTS OF ROUNDING OR GROUPING DATA

## FIGURES

This chapter was prepared by
Churchill Eisenhart, National Bureau of Standards

# CHAPTER 4

## EFFECTS OF ROUNDING OR GROUPING DATA

### 1. INTRODUCTION

A statistician must of necessity work with rounded numbers, at least to the extent that his data are recorded values of continuous variables, that is, measurements. Enumeration data may be rounded also, to tens, hundreds, thousands, or millions, for convenience in handling. The more coarsely the data are rounded, the less accurately do they represent the facts observed, and very coarse rounding may render them useless as a basis for conclusions about the phenomena concerned. This is self-evident; but it is not always appreciated that rounding considerably less coarse than that which renders the data useless may, nevertheless, greatly impair or even invalidate the analysis of the data.

William Sealy Gosset, better known as "Student," has given a most penetrating discussion of the effects of coarse rounding or grouping of data on their statistical properties.* Referring to certain data, relating to height and left middle finger measurements of 3,000 criminals, which he had used for an experimental test of his theoretical results, he wrote:

Another thing which interferes with the comparison is the comparatively large groups in which the observations occur. The heights are arranged in 1-inch groups, the standard deviation being only 2.54 inches, while the finger lengths were originally grouped in millimeters, but unfortunately I did not at the time see the importance of having a smaller unit and condensed them into 2-milli-meter groups, in terms of which the standard deviation is 2.74.

Several curious results follow from taking samples of 4 from material disposed in such wide groups. The following points may be noticed:

(1) The means only occur as multiples of 0.25.

(2) The standard deviations† occur as the square roots of the following types of numbers: $n$, $n + 0.19$, $n + 0.25$, $n + 0.50$, $n + 0.69$, $2n + 0.75$.

(3) A standard deviation belonging to one of these groups can only be associated with a mean of a particular kind; thus a standard deviation of $\sqrt{2}$ can only occur if the mean differs by a whole number from the group we take as origin, while $\sqrt{1.69}$ will only occur when the mean is at $n \pm 0.25$.

(4) All the four individuals of the sample will occasionally come from the same group, giving a zero value for the standard deviation. Now this leads to an infinite value of $z$‡ and is clearly due to too wide a grouping, for although two

---

* STUDENT, "The Probable Error of a Mean," Sec. IV. (For full citations, see Sec. 7.)

† Here Student uses the term "standard deviation" to refer to the root mean square of the deviations of the four observations in a sample from their mean.

‡ Here Student uses $z$ to refer to the deviation of the sample mean from the popula-

men may have the same height when measured by inches, yet the finer the measurements the more seldom will they be identical, till finally the chance that four men will have *exactly* the same height is infinitely small.   If we had smaller grouping the zero values of the standard deviation might be expected to increase, and a similar consideration will show that the smaller values of the standard deviation would also be likely to increase, such as 0.436, when 3 fall in one group and 1 in an adjacent group, or 0.50 when 2 fall in two adjacent groups.   On the other hand, when the individuals of the sample lie far apart, the argument of Sheppard's correction will apply, the real value of the standard deviation being more likely to be smaller than that found owing to the frequency in any group being greater on the side nearer the mode.

These two effects of grouping will tend to neutralize each other in their effect on the mean value of the standard deviation but both will increase the variability.

Accordingly, we find that the mean value of the standard deviation is quite close to that calculated, while in each case the variability is sensibly greater. The fit of the curve is not good, both for this reason and because the frequency is not evenly distributed owing to effects (2) and (3) of grouping.   On the other hand, the fit of the curve giving the frequency of $z$ is very good, and as that is the only practical point the comparison may be considered satisfactory.

The Statistical Research Group received various sets of data that exhibited curious statistical properties.   In some instances it was found that these properties were attributable to the coarse rounding employed in recording the data.   In at least one instance the peculiarities were attributable to the fact that a fraction of the data had been recorded with half the precision of the remainder, owing (in part) to intrinsic differences in the precisions of the two machines used to obtain the measurements. Occasionally the Group was asked to make recommendations with regard to the recording of data, since its investigations had shown that in many cases the trouble arose in connection with recording the data and not in the actual reading of the scales.   As a consequence of difficulties encountered in analysing coarsely rounded data and of requests for advice on rounding in general, a study of rounding was undertaken, principally from the viewpoint of its effect on the standard procedures of statistical analysis.   The results of this study are summarized in this chapter.   Much of the material presented here can be found in the literature of statistics, but it is widely scattered.

## 2. ROUNDED NUMBERS AND SCALE READINGS

Rounded numbers enter the practical man's world through two different doorways.   First, the measurements of a scientist or engineer are rounded numbers, since all measuring equipment is of limited accuracy. Second, data are often recorded to less accuracy than is attainable with the available equipment, the accuracy preserved being more easily

tion mean, divided by the standard deviation of the sample, the latter defined as above (see Sec. 4.4 below).

attained and "adequate for the purposes at hand"; or they are rounded in computing, sometimes to simplify arithmetical calculations, as when 0.40 is used in place of 51/128 (0.3984375 is exact), and sometimes because it cannot be avoided, as when 3.1416 is used in place of $\pi$.*

In the case of the computer's reduction of exact numbers to rounded numbers in decimal form, assuming that he has rounded them correctly in accordance with the rule given in the next section, it can be stated that each is accurate within $\frac{1}{2}$ unit in the last decimal place retained. In the case of the scientist's or the engineer's measurements this may not be true, and yet the last figures retained may contain worth-while information, in the sense that when the last figures are retained the data are more accurate than when the last figures are omitted entirely or are used as a basis for rounding the results to one less place. In such cases, if the measurements are all of comparable accuracy, there will exist a "unit" of measurement with respect to which the uncertainty of a given measurement is $\frac{1}{2}$ a unit.

Consider, for example, the case of an experimenter taking "readings" on a scale with graduations every $\frac{1}{8}$ in. He can read his determination to the nearest $\frac{1}{8}$ in. directly and may then visually estimate it to the nearest $\frac{1}{16}$ in., obtaining, say, a reading of $3\frac{7}{16}$ in. If he writes this out in decimal form it becomes 3.4375 in., and represents some number between 3.40625 in. ($3\frac{13}{32}$ in.) and 3.46875 in. ($3\frac{15}{32}$ in.), and this is accurate only to within $\pm 0.03125$ in. and not to within 0.00005 in., as the number of decimal places retained might suggest. He may decide, therefore, to report the value as 3.44, which is accurate only to within $+0.02875$ and $-0.03375$, but which will be closer to the true value than is 3.4, unless the true value lies between 3.40625 and 3.42000. When so reported, however, his readings will not give a clear indication of their accuracy. If an unambiguous indication of their accuracy is important, he should report his readings in units of $\frac{1}{8}$ in., in which event they will be accurate to within plus or minus $\frac{1}{2}$ unit.

### 3. THE ROUNDING RULE AND ITS PROPERTIES

When a number expressed in the decimal system is to be rounded to a specific number of significant figures,† the rounding should always be carried out in accordance with the following rule:

* The mathematician has devised convenient symbols for certain very practical quantities, such as the diagonal of a square of unit side ($\sqrt{2}$) and the ratio of the circumference of a circle to its diameter ($\pi$), which cannot be expressed *exactly* as common fractions or as decimal numbers with a finite number of decimals. By using the mathematician's symbols for these quantities, the practical man can usually keep his algebraic manipulations simple, but sooner or later there comes a point at which a numerical answer is needed, and here it becomes necessary to replace $\pi$ (3.14159 26535 89793 23846+) with a common fraction, such as 355/113 (which is correct to 6 decimal places), or with a decimal approximation such as 3.141593.

† A *significant figure* of a number (expressed in the decimal system of notation)

*Rounding Rule:*

*a.* If the first digit to be dropped is less than 5, the last digit retained shall be left unchanged.

*b.* If the first digit to be dropped is greater than 5, or is a 5 followed by digits greater than 0, the last digit retained shall be increased by 1.

*c.* If the first digit to be dropped is a 5 alone, or a 5 followed by 0's only, the last digit retained shall be rounded to the nearest even number. Thus, rounding to 2 decimal places,

$$\pi \doteq 3.141593 \text{ yields } 3.14$$
$$e \doteq 2.718282 \text{ yields } 2.72$$
$$1/22 \doteq 0.045455 \text{ yields } 0.05$$
$$1/200 \doteq 0.005000 \text{ yields } 0.00$$
$$1/198 \doteq 0.005051 \text{ yields } 0.01$$
$$10/2,247 \doteq 0.004450 \text{ yields } 0.00$$

It is clear that to every number that is to be "rounded down" in accordance with part (*a*) of the rule there corresponds a number that is to be "rounded up" in accordance with part (*b*). Parts (*a*) and (*b*) thus establish a balance between rounding down and rounding up for all numbers other than those that are exactly 5 in the first place to be dropped. Part (*c*) establishes a similar balance between rounding down and rounding up for numbers that are exactly 5 in the first place to be dropped, by specifying that the rounding shall be up when the preceding digit is 1, 3, 5, 7, and 9; down, when it is 0, 2, 4, 6, 8. Part (*c*) is arbitrary, since rounding to the nearest odd number would establish a balance in these cases equally well, being equivalent to merely interchanging the words up and down in the preceding sentence. However, uniformity of procedure is important for checking, and consistency of published tables is desirable, so part (*c*), being well established, should be adhered to.

A question that arises naturally at this point is: Will repeated applications of the rule (dropping digits singly, or in pairs, or otherwise) always lead to the same result, when the rounding is stopped at a specified place, as a single rounding to that place? The answer, unfortunately, is No. The fifth example in the list above is a case in point. If 0.005051 is rounded to 2 decimal places in accordance with the rule in almost any manner involving more than one step, the result is 0.00, which is "incorrect." The third example may be used to illustrate the fact that the number of places to which a number is originally tabulated may affect the outcome of subsequent rounding: $1/22 = 0.0\ 45\ 45\ 45\ 45$ *ad infinitum;*

is any digit that contributes to the specification of the magnitude of the number aside from serving to locate the decimal point. For example, 345 and 0.000345 each have three significant figures; 34.05 and 34.50 (as distinct from 34.5) each have four significant figures; 34500 has 3, 4, or 5 according as it means $3.45 \times 10^5$, $3.450 \times 10^5$, or $3.4500 \times 10^5$, respectively.

correct to 6 decimal places it equals 0.045 455, the value given in the table above; if this latter value is rounded to 5 decimals in accordance with the rule, the result is 0.045 46; however ½₂ equals 0.045 45 correct to 5 decimals, the result that will be obtained by rounding to 5 decimals when the value originally tabulated is correct to 7 decimals or more.

It is clear from the foregoing that, when an $n$-place table is prepared by rounding the entries given in an $(n + k)$-place table, some of the entries may not be correct in the $n$th place. Thus, in a table of $\log_e N$ to 5 decimal places, the 5-place value given for $\log_e 1947 = 7.57404\ 50054$ will be incorrect in the fifth place if obtained by rounding from less than an 8-place table with entries correct in the eighth place. Final 5's can, however, be marked to show whether further rounding should be up or down; see, for example, Tables 1.1 and 1.5 of Chap. 1, in which each final 5 is italicized if rounding should not be in accordance with part (c) of the rule.

It may be asked whether there exists a rounding rule that will always lead to the same result at a specified place, irrespective of the number of applications of the rule employed in rounding to that place. A procedure suggested by Cayley* possesses this advantage. According to this procedure, the figure in the last place to be retained is not adjusted at all—which is equivalent to rounding down irrespective of the digits that follow—with the result that the number recorded either is exact or errs in defect. If, when a recorded number of this type is used, a final 5 is affixed to it, then the error will lie between plus and minus ½ unit in the last place recorded, as is the case when the number recorded is rounded in a single step from its exact value in accordance with the Rounding Rule. For example, $\pi$ would be recorded to 4 decimals as 3.1415, and would be taken to be 3.14155 in further calculations. Whereas Cayley's procedure has the advantages that it is simple (since numbers are always rounded down), that it can be mechanized for purposes of machine computation, and that it yields an $n$-place recorded number that is independent of the number of rounding steps involved in reaching it from the exact value, it has the disadvantage that it requires the use of $(n + 1)$-place numbers in further computations in order to attain the same degree of accuracy that would be obtained by using $n$-place numbers rounded in accordance with the Rounding Rule.

## 4. EFFECTS OF ROUNDING ON STATISTICAL ANALYSES OF LARGE NUMBERS OF OBSERVATIONS

### 4.1. Nature of the Effects

Frequently a statistician analyzes measurements recorded to several decimal places more than can possibly have any practical significance.

* CAYLEY, letter to J. W. L. Glaisher.

The extra decimal places do not occasion any great concern, for nothing has been done that cannot be undone. Recording the extra digits and rounding the measurements are merely unnecessary nuisances. On the other hand, measurements are sometimes recorded to the nearest 5, 10, 20, 50, or 100 units, or still more coarsely, when they could, and should, have been recorded to the nearest unit, or even more accurately. Thus, if duplicate measurements taken under essentially the same conditions tend to differ on the average by about 50 units, research workers often record the measurements to the nearest 10 or 20 units; sometimes only to the nearest 50 units. If measurements taken under essentially the same conditions tend to be distributed approximately normally about a "true value" with a standard deviation of $\sigma$, these roundings correspond to intervals with widths roughly $0.23\sigma$, $0.45\sigma$, and $1.1\sigma$, respectively.* Statisticians, on the other hand, recommend that the width of the rounding interval be less than one-third, or, better, less than one-fourth the standard deviation of random sampling.†

Basically, the choice of the size of a rounding or grouping interval depends on the answers to the following three questions:–

a. If the observations $x_1, x_2, \ldots, x_N$ are to be regarded as independent values of a random variable $x$, which, in the population from which the data are drawn, is distributed about a mean of $\mu$ with a standard deviation $\sigma$, then what is the effect of the rounding (grouping) on the sample estimates of $\mu$ and $\sigma$, defined by

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N} \tag{1}$$

and

$$s = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N - 1}} \tag{2}$$

b. What is the effect of the rounding on confidence limits for $\mu$ and $\sigma$ based on $\bar{x}$ and $s$?

c. What is the effect of the rounding on statistical tests of hypotheses about $\mu$ and $\sigma$ based on $\bar{x}$ and $s$?

In the present section, answers to these questions will be given for the case when the number of observations $N$ is large. In Sec. 5 the case where $N$ is small, perhaps only 2 or 3, is discussed; the results given there

---

* The average difference between duplicate measurements is, under the assumptions stated, $2\sigma/\sqrt{\pi} = 1.12838\sigma$. If this average difference is 50, a grouping interval of 10 corresponds with $1.12838\sigma/5 = 0.226\sigma$, and similarly for a grouping interval of 20.

† See, for example, FISHER, *Statistical Methods*, Sec. 13.

throw light on what values of $N$ may be regarded as large and what values must be regarded as small.

## 4.2. Sheppard's Corrections

In order to answer the first question, consider the relation between the $k$th (population) moment

$$\mu'_k = E(x^k) = \int_{-\infty}^{+\infty} x^k p(x)\, dx \qquad (k = 1, 2, 3, 4, \cdots) \qquad (3)$$

of a continuous random variable distributed according to the probability density function* $p(x)$, and the $k$th moment

$$\nu'_k = \sum_{i=-\infty}^{+\infty} x_i^k \int_{x_i - \frac{w}{2}}^{x_i + \frac{w}{2}} p(t)\, dt \qquad (k = 1, 2, 3, 4, \cdots) \qquad (4)$$

of the histogram formed by grouping the values of $x$ into classes of width $w$ with mid-points $x_i$, $(i = 0, \pm 1, \pm 2, \cdots)$. The $\mu'_k$ are, by definition, the moments (about the origin) of $x$ in the population, in which $x$ is a continuous variable. On the other hand, since it is not possible either to read or to record with perfect accuracy all values of a continuous variable, recorded values of $x$ will generally be rounded values, any value of $x$ between $x_i - (w/2)$ and $x_i + (w/2)$ being recorded as $x_i$. Hence, in a sense, the $\nu'_k$ are *realistic* moments of $x$ and the $\mu'_k$ are *ideal* moments.

If, first, $p(x)$ can be expanded in a Taylor series throughout the range for which $p(x) > 0$, second, $x^k p^{(2n)}(x)$ is finite and continuous throughout this range ($k = 1, 2, 3, 4, \cdots$ ; $n = 1, 2, 3, 4, \cdots$), and third, $p(x)$ and $p^{(n)}(x)$ ($n = 1, 2, \cdots$) vanish at the limits of this range, then the $\nu'_k$ and the $\mu'_k$ are related as follows:

$$\nu'_1 = \mu'_1$$

$$\nu'_2 = \mu'_2 + \frac{w^2}{12}$$

$$\nu'_3 = \mu'_3 + \left(\frac{w^2}{4}\right)\mu'_1$$

$$\nu'_4 = \mu'_4 + \left(\frac{w^2}{2}\right)\mu'_2 + \frac{w^4}{80} \qquad (5)$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\nu'_k = \sum_{i=0}^{\infty}\left[C^k_{2i}\left(\frac{w}{2}\right)^{2i}(2i + 1)\right]\mu'_{k-2i}$$

where $C^k_r = k!/[r!(k - r)!]$ for $0 \le r \le k$, and $C^k_r = 0$ for $r > k$.†

---

* That is, for any value of $c$, the integral $P(c) = \int_{-\infty}^{c} p(x)\, dx$ equals the probability that $x$ is less than $c$, $0 \le P(c) \le 1$ for all values of $c$, $P(-\infty) = 0$, and $P(+\infty) = 1$.

† SHEPPARD, "Most Probable Values of Frequency Constants." The corresponding case of two variables has been considered by Baten, "Corrections for the Moments," and of several variables by Wold, "Sheppard's Correction Formulae."

The three conditions regarding $p(x)$ imply that it is not only continuous over the range for which $p(x) > 0$ but tapers off strongly to zero at the ends of this range; they thus exclude probability curves with abrupt ends, such as a uniform distribution over $-w/2$ to $+w/2$, and $U$-shaped or $J$-shaped curves. The three conditions are necessary and sufficient for the relations (5) to hold exactly. Relations analogous to (5), with additional terms, have been obtained for cases excluded by the above assumptions.*

The relations (5) can be derived in a somewhat more general manner that shows that they are valid *on the average* whenever the moments involved exist, provided that the rounding lattice is imposed at random on the underlying distribution.† Suppose the $x$ axis is first divided in a definite manner into an infinite number of class intervals of width $w/m$, where $m$ is some integer greater than 1. Given a probability density function $p(x)$, the histogram corresponding to this grouping of the values of $x$ can always be determined from

$$p(x_j) = \int_{x_j - \left(\frac{w}{2m}\right)}^{x_j + \left(\frac{w}{2m}\right)} p(t)\, dt \qquad (j = 0, \pm 1, \pm 2, \cdots) \qquad (6)$$

where $x_j$ is the mid-point of the $j$th subdivision, since the integrals involved will always exist when $p(x)$ is a probability density function. The moments about the origin of this basic histogram, $\mu_k''$, say, can be determined from the summation analogous to (4), so far as they exist (that is, are finite). Corresponding to the basic histogram determined by (6), $m$ new histograms can be formed by combining the successive subclasses of width $w/m$ into subclasses of width $w$, the intervals containing $x_0, x_1, \ldots, x_{m-1}$ being combined in one instance, the intervals containing $x_1, x_2, \ldots, x_m$ being combined in the next instance, and so forth. Corresponding to each of these histograms, moments about the origin, $\nu_k'(s)$, $s = 1, 2, \cdots, m$, can be determined from summations analogous to (4), so far as they exist. Let $\bar{\nu}_k'$ denote the average of the values of $\nu_k'(s)$, $s = 1, 2, \cdots, m$, then the $\bar{\nu}_k'$ and the $\mu_k''$ are related as follows, in so far as they exist:

$$\bar{\nu}_1' = \mu_1''$$
$$\bar{\nu}_2' = \mu_2'' + \left[1 - \frac{1}{m^2}\right] \cdot \frac{w^2}{12}$$
$$\bar{\nu}_3' = \mu_3'' + \left[1 - \frac{1}{m^2}\right] \cdot \frac{w^2}{4} \cdot \mu_1'' \qquad (7)$$
$$\bar{\nu}_4' = \mu_4'' + \left[1 - \frac{1}{m^2}\right] \cdot \frac{w^2}{2} \cdot \mu_2'' + \left[1 - \frac{1}{m^2}\right] \cdot \left[3 - \frac{7}{m^2}\right] \cdot \frac{w^4}{240}$$

* See papers by Elderton, Pairman and Pearson, Pearse, and Pearson listed in Sec. 7 below.

† See the papers by Abernathy, Craig, and Carver listed in Sec. 7.

and so forth.    Letting $m \to \infty$, the right-hand sides of these equations tend to the corresponding expressions in equations (5), since $\mu_k'' \to \mu_k'$ as $m \to \infty$.    On the other hand, the $\bar{\nu}_k'$ retain their meanings as average values of the $\nu_k'$.    Accordingly, whatever the nature of the distribution of $x$, equations (5) hold *on the average*, that is, they give the average relationships between the $\nu_k'$ and the $\mu_k'$, provided these are finite, for an infinite series of trials in which a rounding lattice composed of sections of width $w$ is imposed at random on the underlying probability distribution.

The results given above all refer to moments about a fixed point taken as the origin, and the values of the moments involved will depend upon the choice of the origin.    Thus, if $x$ denotes temperature and is measured in degrees Fahrenheit, the values of the $\mu_k'$ and $\nu_k'$ will be different from those obtained when $x$ is measured in degrees centigrade.    The results obtained will be the same, however, if the origin is chosen at the mean of the distribution of $x$.    Dropping the primes to signify moments about the mean of $x$, relations (5) correspond with

$$\nu_1 = \mu_1 = 0$$
$$\nu_2 = \mu_2 + \frac{w^2}{12}$$
$$\nu_3 = \mu_3$$
$$\nu_4 = \mu_4 + \frac{w^2}{2} \cdot \mu_2 + \frac{w^4}{80}$$

(8)

and so forth.

Comparison of the second equation of (8) with (22) shows that the variance (second moment about the mean) of a rounded value of $x$, which may be denoted by $x_R$ for convenience, equals the variance of $x$, namely $\mu_2$, plus a variance $w^2/12$, which equals the variance of a rounding error $\epsilon$, which is distributed uniformly between $-w/2$ and $+w/2$.    This suggests that we may write

$$x_R = x + \epsilon$$

(9)

where $x_R$ denotes a value of $x$ rounded to within $\pm w/2$, and $\epsilon$ is a random variable independently distributed with uniform probability density between $-w/2$ and $+w/2$.    That the representation (9) is strictly valid under the three conditions stated above, and that it is valid *on the average*, in the above sense, under any conditions for which the moments exist, has been demonstrated rigorously.[*]

Since it is usually desired to determine the $\mu_k$ from the $\nu_k$, equations (8) are customarily solved for the $\mu_k$ in terms of the $\nu_k$ and powers of $w$, and the relations given as

[*] See papers by Craig and by Langdon and Ore, listed in Sec. 7.

$$\mu_1 = \nu_1 = 0$$

$$\mu_2 = \nu_2 - \frac{w^2}{12} \qquad (10)$$

$$\mu_3 = \nu_3$$

$$\mu_4 = \nu_4 - \frac{w^2}{2} \cdot \nu_2 + \frac{7w^4}{240}$$

and so forth.*

These adjustments of the $\nu_k$ to obtain the $\mu_k$ are customarily called Sheppard's corrections for grouping. As has been indicated above, they are valid as average adjustments whenever the moments involved exist, provided that the rounding lattice is imposed *at random* on the underlying distribution. This latter provision means, for example, that when values of a variable that can take only nonnegative values are to be rounded, zero shall not inevitably be adopted as the lower extremity of the lowest class—indeed, the mid-point of the lowest class may turn out to be a small negative number. If the rounding lattice is not imposed at random when Sheppard's original assumptions are not fulfilled, then the foregoing adjustments may be wildly wrong and may even adjust in the wrong direction.†

### 4.3. Comparison of the Distribution of the Mean of $N$ Rounded Observations and the Distribution of the Mean of $N$ Unrounded Observations, When $N$ Is Large

Sheppard's corrections make possible a concise comparison of the distribution of the mean of $N$ rounded observations and the distribution of the mean of $N$ unrounded observations, when $N$ is large. Continuing the notation adopted in connection with (9), let

$$\bar{x}_R = \sum_{i=1}^{N} \frac{x_{Ri}}{N} \qquad (11)$$

where $x_{Ri}$ is the rounded value of $x_i$, the $i$th of $N$ independent values of a random variable $x$, corresponding to a rounding lattice with classes of

---

* The corresponding relations for $\mu_5$ to $\mu_8$ are given in Rietz, *Handbook*, p. 94. They can be obtained by letting $m \to \infty$ in the expressions for $\mu_5$ to $\mu_8$ given by Craig. Letting $\lambda_k$ and $L_k$ ($k = 1, 2, \cdots$) denote the semi-invariants of the unrounded variable $x$ and the rounded variable $x_R$, respectively, the corresponding relations become simply

$$\lambda_{2m} = L_{2m} + (-1)^m B_m \frac{w^{2m}}{2m}$$

$$\lambda_{2m+1} = L_{2m+1}$$

where $B_m$ is the $m$th Bernoulli number. Thus, $\lambda_4 = L_4 + w^4/120$, $\lambda_6 = L_6 - w^6/252$, $\lambda_8 = L_8 + w^8/240$, and so forth. For further details, see the papers by Craig, Langdon and Ore, and Wold.

† The reader who wishes to delve further into the mathematical background of Sheppard's corrections is urged to read Kendall's paper.

width $w$ (located at random on the $x$ axis, if Sheppard's assumptions are not fulfilled). Furthermore, let

the mean of $x$ in the population $\qquad = \mu'_{1:x}$
the variance of $x$ in the population $\qquad = \sigma_x^2 = \mu_{2:x}$ (12)
the $k$th moment of $x$ about its population mean $= \mu_{k:x}$
$$(k = 1, 2, \cdots)$$

and let

$$w = a \cdot \sigma_x \qquad (13)$$

Table 4.1 presents a comparison of the distribution of $\bar{x}_R$ in regard to mean, variance, skewness, and kurtosis with the distribution of $x$ and with a normal distribution of mean $\mu'_1$ and variance $\sigma^2$.

From a study of Table 4.1 it is apparent that

(1) $\bar{x}$ and $\bar{x}_R$ are each unbiased estimators of the population mean of $x$.

(2) The variance of $\bar{x}_R$ is $(1 + a^2/12)$ times the variance of $\bar{x}$.

(3) The skewness ($\gamma_1 = + \sqrt{\beta_1}$ of Pearson's notation) of the distribution of $\bar{x}_R$ never exceeds that of the distribution of $\bar{x}$.

(4) The kurtosis ($\gamma_2 = \beta_2 - 3$ of Pearson's notation) of the distribution of $\bar{x}_R$ is always less than that of the distribution of $\bar{x}$.

(5) When $N$ is large, $\bar{x}$ and $\bar{x}_R$ are approximately normally distributed about the population mean of $x$, $\mu'_{1:x}$.

The case in which the ungrouped variable $x$ is normally distributed is of special interest. In this case $\bar{x}$ is likewise normally distributed, whereas the distribution of $\bar{x}_R$ is symmetric but has slightly less kurtosis than a normal distribution, the deficiency being $(a^4/120N)/(1 + a^2/12)^2$. From the Gram-Charlier series for the "generalized normal curve,"[*] it is seen that this deficiency in kurtosis implies that the distribution of $\bar{x}_R$ has slightly less area in the "central region," say, between $\pm 0.75 \sigma_{\bar{x}_R}$, slightly less also in the "tails," say, beyond $\pm 2.3 \sigma_{\bar{x}_R}$, and slightly more between these points. For any practical rounding, say, $a = w/\sigma < 1$, this effect will surely be negligible. R. A. Fisher has shown[†] that the maximum likelihood estimate $\hat{\mu}$ of $\mu'_{1:x}$ from $N$ independent rounded values $x_R$ of a normally distributed variable $x$ has a variance of $(1 + a^2/12 - a^6/2{,}880 + \cdots) \cdot \sigma_x^2/N$. The loss of precision when $\bar{x}_R$ is used instead of $\hat{\mu}$ is, therefore, only $a^6/2{,}880$, which is clearly negligible for any practical rounding.

In conclusion

a. $\bar{x}_R$ is an unbiased estimator of the population mean of $x$, $\mu'_{1:x}$, whatever the value of $N$.

b. When $N$ is large and the values of $\sigma_x^2$ and $w$ are known, so that $\sigma^2_{\bar{x}_R} = (\sigma_x^2 + w^2/12)/N$ can be evaluated, confidence limits for, and

[*] See, for instance, RIETZ, *Handbook*, p. 114.
[†] FISHER, "Mathematical Foundations," p. 362.

statistical tests of hypotheses regarding the value of $\mu'_{1:x}$ carried out with $(\bar{x}_R - \mu'_{1:x})/\sigma_{\bar{x}_R}$ regarded as a normal deviate will be approximately correct—provided the true distribution of $x$ has "high order contact" (that is, Sheppard's assumptions are valid), or the rounding lattice is located at random with respect to the admissible values of $x$, or both.

TABLE 4.1

COMPARISON OF THE MEAN, VARIANCE, SKEWNESS, AND KURTOSIS OF THE DISTRIBUTION OF THE AVERAGE OF $N$ INDEPENDENT ROUNDED VALUES OF A RANDOM VARIABLE WITH THE CORRESPONDING CHARACTERISTICS OF THE DISTRIBUTION OF THE UNROUNDED VARIABLE AND OF A NORMAL (GAUSSIAN) DISTRIBUTION WHEN THE WIDTH OF THE ROUNDING INTERVAL IS $a$ STANDARD DEVIATION UNITS ($w = a \cdot \sigma$)

| Characteristic | Distribution | | |
|---|---|---|---|
| | Unrounded variable, $x$ | Average of $N$ rounded values, $\bar{x}_R$ | Normal |
| Mean.......................... | $\mu'_{1:x}$ | $\mu'_{1:x}$ | $\mu'_1$ |
| Variance....................... | $\sigma_x^2$ | $\dfrac{\left(1 + \dfrac{a^2}{12}\right)\sigma_x^2}{N}$ | $\sigma^2$ |
| Skewness ($\gamma_1$)................ | $\gamma_{1:x} = \dfrac{\mu_{3:x}}{\sigma_x^3}$ | $\dfrac{\gamma_{1:x}}{\sqrt{N\left(1 + \dfrac{a^2}{12}\right)^3}}$ | $0$ |
| Kurtosis ($\gamma_2$)................. | $\gamma_{2:x} = \dfrac{\mu_{4:x}}{\sigma^4} - 3$ | $\dfrac{\gamma_{2:x} - \dfrac{a^4}{120}}{N\left(1 + \dfrac{a^2}{12}\right)^2}$ | $0$ |

## 4.4. Comparison of the Distribution of $s^2$ for Rounded and Unrounded Observations, When the Number of Observations Is Large

Let

$$s^2_{x_R} = \sum_{i=1}^{N} \frac{(x_{Ri} - \bar{x}_R)^2}{N - 1} \tag{14}$$

where $x_{Ri}$ ($i = 1, 2, \cdots, N$) and $\bar{x}_R$ are as in (11). In Table 4.2 the mean of $s^2_{x_R}$ and the ratio of the variance of $s^2_{x_R}$ to the square of its mean are compared with the mean and corresponding ratio for $s^2_x$ under two sets of conditions: first, when $x$ has any distribution for which the first four moments exist; second, when $x$ is normally distributed.

It is evident from a study of Table 4.2 that the ratio

$$(\text{variance of } s^2)/(\text{mean of } s^2)^2$$

is always less in the case of $s^2_{x_R}$ than in the case of $s^2_x$. The difference is slight, however, and for any reasonable rounding, say, $a = w/\sigma < 1$, will generally be negligible. Therefore

(1) When $N$ is large, $(N - 1)s^2_{x_R}/(\sigma^2_x + w^2/12)$ and $(N - 1)s^2_x/\sigma^2_x$ will be distributed as $\chi^2$ with $N - 1$ degrees of freedom to about the same order of approximation.

TABLE 4.2

THE MEAN OF $s^2_{x_R}$ AND OF THE RATIO OF THE VARIANCE OF $s^2_{x_R}$ TO THE SQUARE OF THE MEAN COMPARED WITH THE CORRESPONDING QUANTITIES FOR $s^2_x$, WHEN THE DISTRIBUTION OF $x$ IS ARBITRARY* AND WHEN THE DISTRIBUTION OF $x$ IS NORMAL

$$\left(a = \frac{w}{\sigma_x}\right)$$

| Distribution of $x$ | Arbitrary | | Normal |
|---|---|---|---|
| Statistic | $\varepsilon^2_x$ | $s^2_{x_R}$ | $s^2_x$ |
| Mean | $\sigma^2_x$ | $\left(1 + \dfrac{a^2}{12}\right) \cdot \sigma^2_x$ | $\sigma^2_x$ |
| $\dfrac{\text{Variance}}{(\text{Mean})^2}$ | $\dfrac{2}{N - 1} + \dfrac{\gamma_{2:x}}{N}$ | $\dfrac{2}{N - 1} + \dfrac{\gamma_{2:x} - \dfrac{a^4}{120}}{N}$ | $\dfrac{2}{N - 1}$ |

* Subject to the restriction that the first four moments of $x$ about its population mean exist.

(2) The quantity $s^2_{x_R} - w^2/12$ is an unbiased estimator of the variance of $x$ in the population $\sigma^2_x$, whatever the value of $N$, provided the distribution of $x$ has high order contact (that is, Sheppard's three assumptions, stated in Sec. 4.2, are fulfilled), or the rounding lattice is located at random with respect to the admissible values of $x$, or both.*

It may be well to point out here that the ratio

$$(N - 1)(s^2_{x_R} - w^2/12)/\sigma^2_x$$

should not be used as a substitute for the ratio given in (1) unless $a = w/\sigma$ is quite small, since the variance of this ratio exceeds the variance of the ratio given in (1) by a factor of $(1 + a^2/12)^2$, which will generally render

* When $N$ is small, the mean and variance/(mean)² relations shown in Table 4.2 are still valid, but the discontinuities of the distribution of $s^2_{x_R}$ are such that graduation of the distribution of $s^2_{x_R}$ by a continuous distribution is virtually impossible (see Sec. 5.1).

Kendall has remarked that (2) gives the correct estimator of $\sigma^2_x$ if the rounding of the values of $x$ is determined before the sample is drawn. If, on the other hand, the observations $x_1, x_2, \ldots, x_N$ are drawn first and then rounded, he suggests that $s^2_{x_R} - w^2/[12(N - 1)]$ be used to estimate $\sigma^2_x$. When $N$ is large the difference is clearly negligible from a practical standpoint but is of some theoretical interest.

the $\chi^2$ approximation less accurate. If the adjusted numerator and the denominator $\sigma_x^2 + w^2/12$ are used, then the variance of the resulting ratio will be the same as the variance of the ratio given in (1), but the mean will be deficient by a factor of $1/(1 + a^2/12)$. If $s_{x_R}^2$ is used in the numerator and $\sigma_x^2$ in the denominator, the mean will be excessive by a factor of $(1 + a^2/12)$ and the variance by a factor of $(1 + a^2/12)^2$. Consequently,

(3) Determinations of confidence limits for, and statistical tests of hypotheses regarding, the value of $\sigma_x^2$ should be based on $s_{x_R}^2$ as implied by (1) and will be valid to the extent that the conditions and approximations thereof obtain.

In conclusion, it should be remarked that the foregoing discussion can be carried over to the case where $s_{x_R}^2$ has been obtained from an analysis of variance of values of $x$ that have been rounded to within $\pm w/2$ and has $n$ degrees of freedom. The only change required is replacement of $N - 1$ by $n$ throughout.

## 4.5. Effect of Rounding on the Distribution of Student's $t$

Let

$$t_x = \frac{\sqrt{N}\,(\bar{x} - \mu'_{1:x})}{s_x} \tag{15}$$

and

$$t_{x_R} = \frac{\sqrt{N}\,(\bar{x}_R - \mu'_{1:x})}{s_{x_R}} \tag{16}$$

where $\bar{x}$ and $\bar{x}_R$ are as in (1) and (11), $s_x$ and $s_{x_R}$ are as in (2) and (14), $\mu'_{1:x}$ is the mean value of $x$ in the population, and $N$ is the number of independent values of $x$ (or of $x_R$) in the sample.

When the $N$ independent observations involved are from a normal distribution, the distribution of $t_x$ is known as "Student's $t$ distribution," in recognition of the fact that the distribution of $z = t/\sqrt{N-1}$ under these conditions was first given by Student in the epochal paper from which the quotation in Sec. 1 of this chapter is taken. In his derivation, Student utilized the independence of $\bar{x}$ and $s$ when computed from independent observations from a normal population, although he himself merely established the fact that the product moment coefficients of correlation between $\bar{x}$ and $s$, and between $\bar{x}^2$ and $s^2$, are zero under these conditions. Independence of $\bar{x}$ and $s$ is actually a necessary condition. A rigorous proof of their independence in random samples from a normal population was communicated to Student by R. A. Fisher in correspondence prior to the First World War; this first saw light, only incidentally, as a preliminary step in Fisher's paper, published in 1915, on the exact sampling distribution of the correlation coefficient.* The first complete

* Fisher, "Distribution of the Correlation Coefficient."

derivation of Student's distribution was given, in terms of $t$, by Fisher in 1926,[*] although he had given a more general definition of $t$ two years earlier.[†]

As Student pointed out in the passage quoted in Sec. 1 of this chapter, there is a correlation between the sample mean and the sample standard deviation in random samples of rounded observations, resulting from the rounding, so that $\bar{x}_R$ and $s_{x_R}$ will not be independently distributed even when $\bar{x}$ and $s_x$ are independent. Nevertheless, Student's experimental results suggest that the distribution of $t_{x_R}$ will closely approximate that of $t_x$ when $N$ is large enough to render negligible the discontinuities of the distribution of $t_{x_R}$. That this is so is evident from a study of the first four moments of $t_{x_R}$ for the case of $x$ having an arbitrary distribution subject only to the condition that its first four moments exist. The first four moments of $t_{x_R}$ can be evaluated readily, under this assumption regarding $x$, from the expressions for the first four moments of $t$ in terms of the semi-invariants of $x$, given by Geary.[‡] The first four moments of $t_{x_R}$ are given in Table 4.3, where they are compared with the first four moments of Student's $t$ distribution, that is, the moments of $t_x$ when $x$ is normally distributed.[§]

It is evident from a study of Table 4.3 that, to terms of order $1/N$, rounding affects only the fourth moment, and that the effect is clearly negligible in the case of samples from a normal population ($\gamma_{1:x} = 0$ and $\gamma_{2:x} = 0$) for any reasonable rounding, say, $a = w/\sigma < 1$. It can be seen from a study of Geary's more extensive formulas that, when the distribution of $x$ is symmetrical but not necessarily normal, the first and third moments of $t_{x_R}$ are zero through terms in $1/N^2$ at least and the variance of $t_{x_R}$ has its normal value through the terms in $1/N$. Indeed, when $x$ is normally distributed, the variance of $t_{x_R}$ has its normal value through terms in $1/N^2$ also. Therefore, it may be stated that

When $N$ is sufficiently large and $w$ is sufficiently small to render the discontinuities of the distribution of $t_{x_R}$ negligible, the distribution of $t_{x_R}$ and $t_x$ will follow the standard $t$ distribution to about the same order of approximation, provided the distribution of $x$ has high order contact, or the rounding lattice is located at random with respect to the admissible values of $x$, or both.

---

[*] FISHER, "Applications of Student's Distribution."

[†] FISHER, "Error Function of Well-Known Statistics."

[‡] GEARY, "Distribution of Student's Ratio."

[§] The moments of $t_{x_R}$ are given to terms of order $1/N$; that is, terms in the $3/2$ and higher powers of $1/N$ have been neglected. Geary's general expressions for the first four moments are correct to terms of order $1/N^2$. The terms of order $1/N^{3/2}$ and $1/N^2$ for $t_{x_R}$ can be evaluated with the aid of the semi-invariants of $x_R$ given in the footnote to expressions (10) above.

TABLE 4.3

First Four Moments of the Distribution of Student's $t$ When $t$ Is Based on a Sample of $N$ Independent Rounded Observations from Any Population with Finite First Four Moments, Compared with the Corresponding Moments of the Standard $t$ Distribution

$$\left(a = \frac{w}{\sigma}\right)$$

| Function | Arbitrary, with finite first four moments | Normal |
|---|---|---|
| $t$ | $t_{z_R}$ | $t_z$ |
| $\mu'_1(t)$ | $-\dfrac{\gamma_{1:z}}{2\sqrt{N}} + 0\left(\dfrac{1}{N^{3/2}}\right)$ | $0$ |
| $\mu'_2(t)$ | $1 + \dfrac{2(1+\gamma_{1:z})}{N} + 0\left(\dfrac{1}{N^2}\right)$ | $\dfrac{N-1}{N-2} = 1 + \dfrac{2}{N} + 0\left(\dfrac{1}{N^2}\right)$ |
| $\mu'_3(t)$ | $-\dfrac{7\gamma_{1:z}}{2\sqrt{N}} + 0\left(\dfrac{1}{N^{3/2}}\right)$ | $0$ |
| $\mu'_4(t)$ | $3\left\{1+\dfrac{6}{N}\left[1+\dfrac{14\gamma_{1:z}^2}{9} - \dfrac{\gamma_{2:z} - \dfrac{a^4}{120}}{9\left(1+\dfrac{a^2}{12}\right)^2}\right]\right\} + 0\left(\dfrac{1}{N^2}\right)$ | $\dfrac{3(N-1)^2}{(N-3)(N-5)} = 3\left(1+\dfrac{6}{N}\right) + 0\left(\dfrac{1}{N^2}\right)$ |

It should be noted that $s^2_{x_R}$, and *not* its adjusted value given in Paragraph (2) of Sec. 4.4, is to be used in the denominator of $t_{x_R}$. If the adjusted value is used, giving

$$t'_{x_R} = \frac{\sqrt{N}\,(\bar{x}_R - \mu'_{1:x})}{\sqrt{s^2_{x_R} - \dfrac{w^2}{12}}} \tag{17}$$

then, since $|t'_{x_R}| > |t_{x_R}|$, it is clear that significant values of $t'_{x_R}$ will arise too frequently. Furthermore, when $s^2_{x_R}$ is less than $w^2/12$, the value of $t'_{x_R}$ will be imaginary, and in such cases some convention regarding the value of $t'_{x_R}$ to be recorded will be needed. In other words, as Fisher has stated, "These [Sheppard's] adjustments should be used for purposes of estimation, but not usually for tests of significance."*

The foregoing discussion carries over to the case where $s^2_{x_R}$ has been obtained by an analysis of variance, or otherwise, from values of $x$ rounded to within $\pm w/2$ and has $n$ degrees of freedom, provided the situation is such that $\bar{x}$ and $s^2_x$ may be regarded as independently distributed. The only change required is to replace $N - 1$ by $n$ throughout.

## 5. EFFECTS OF ROUNDING ON STATISTICAL ANALYSES OF SMALL NUMBERS OF OBSERVATIONS FROM A NORMAL POPULATION

### 5.1. Introduction

Although the results obtained above are independent of the value of $N$, the discussion up to this point has been concerned mainly with cases involving a large number of observations. An exception is the quotation from Student in Sec. 1, which discusses in detail the effects of a coarse rounding on samples of 4 from a nearly normal population. From this passage it is clear that in the case of small samples the discontinuities of the distributions of $\bar{x}_R$ and $s^2_{x_R}$ are the most serious effect of the rounding. For a fixed $N$, the discontinuities of the distributions of $\bar{x}_R$ and $s^2_{x_R}$ may be expected to increase in number (in any finite interval) but to decrease in seriousness as the width of the rounding interval $w$ decreases, and, for a fixed $w$, to increase in number but decrease in seriousness as $N$ increases; hence, it is clear that *largeness* of $N$ has meaning only in conjunction with *smallness* of $w$.

### 5.2. Exact Distributions of $s^2_{x_R}$ in Samples of 2 and 3 from a Normal Population

**5.2.1. Illustrative Distributions.**—In order to throw light on the joint effect of $w$ and $N$ on the distribution of $s^2_{x_R}$ in small samples from a

---

* FISHER, *Statistical Methods*, Sec. D of Appendix to Chap. III.

normal population, exact distributions of $s_{x_R}^2$ in samples of 2 from a normal population have been computed for $w = \sigma$, $\sigma/2$, and $\sigma/4$. These are portrayed in Figs. 4.1, 4.2, and 4.3, where they are compared with

$$N=2, \quad w=\sigma$$
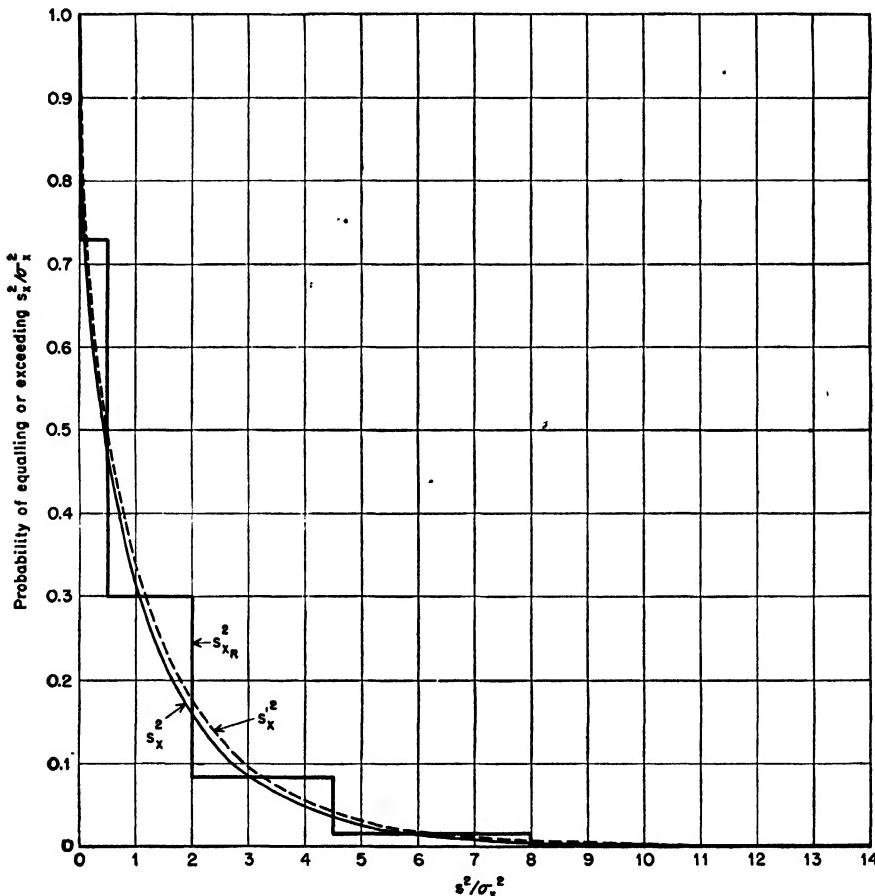


FIG. 4.1.—Distribution of $s_{x_R}^2$ for samples of 2 from a normal population when the width of the rounding interval is one standard deviation, and corresponding continuous distributions of $s_x^2$ and $s_x'^2 = (13/12)s_x^2$.

the distribution of $s_x^2$, which is distributed as $\chi^2\sigma^2$ for 1 degree of freedom. In Fig. 4.1 the distribution of

$$s_x'^2 = \left(1 + \frac{1}{12}\right)s_x^2 \tag{18}$$

is shown also, in view of (1) in Sec. 4.4.

The large steps in the distributions of $s^2_{x_R}$ shown in these figures emphasize the discontinuous character of these distributions. It follows from (1) of Sec. 4.4 that when $N$ is large the distribution of $s'^2_x$ may be expected to graduate the distribution of $s^2_{x_R}$ more closely than will

$$N = 2, \quad w = \sigma/2.$$
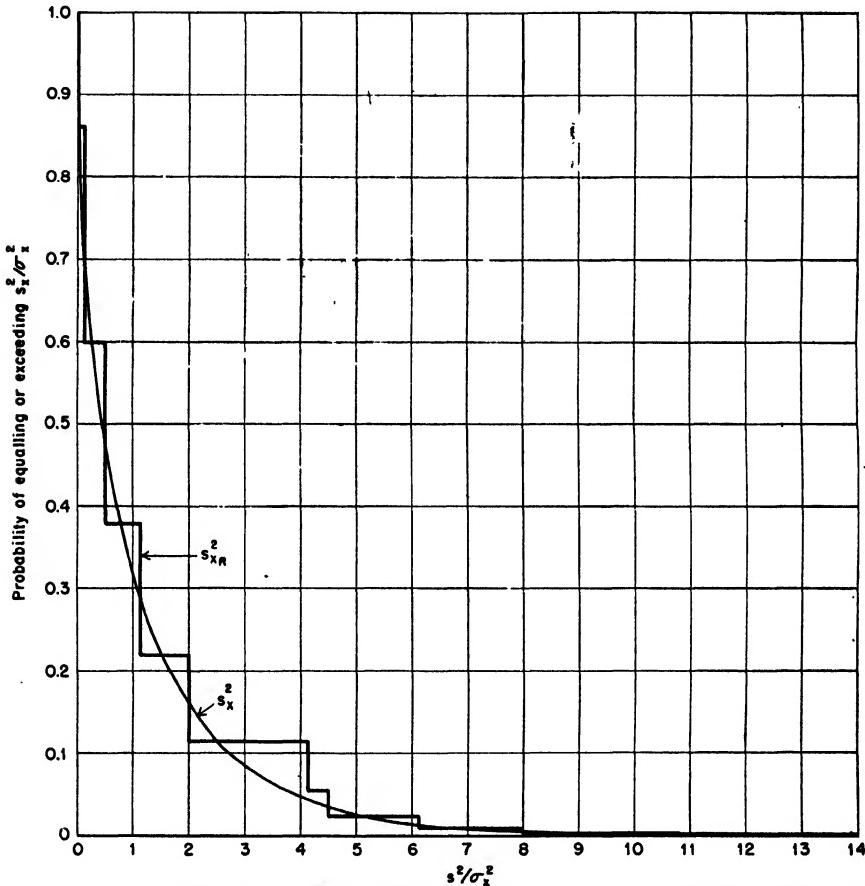


FIG. 4.2.—Distribution of $s^2_{x_R}$ for samples of 2 from a normal population when the width of the rounding interval is one-half standard deviation, and corresponding continuous distribution of $s^2_x$.

the distribution of $s^2_x$. From Fig. 4.1 it is clear that the improvement is negligible when $N = 2$ and $w = \sigma$. Furthermore, it is clearly impossible to find smooth curves that will approximate these step functions closely at all points. In other words, the probability of $s^2_x$, or $s'^2_x$, equaling or exceeding a value $u$ will approximate closely the probability of $s^2_{x_R}$ equaling or exceeding $u$ only for certain values of $u$, principally those that

sampling is from a normal population. These two positions have been chosen as follows:

*Position* I: The mean of $x$ in the population coincides with the boundary between two of the rounding intervals. In the notation of the footnote, this means that if the mean of $x$ in the population is chosen as the origin, so that $\mu'_{1:x} = 0$, then $\theta = -(2i - 1)\pi$, where $i$ is some positive or negative integer, or zero.

TABLE 4.4

EXACT DISTRIBUTIONS OF $s^2_{x_R} = (x_{R_1} - x_{R_2})^2/2$ FOR THE EXTREME POSITIONS OF THE ROUNDING LATTICE WHEN THE WIDTH OF THE ROUNDING INTERVAL IS ONE STANDARD DEVIATION AND $x_1$ AND $x_2$ ARE INDEPENDENT OBSERVATIONS FROM A NORMAL POPULATION

$(N = 2, w = \sigma_x)$

| $u^*$ | Probability that $\dfrac{s^2}{\sigma^2} \geq u$ | |
|---|---|---|
| | Position I | Position II |
| 0 | 1.0000 000 | 1.0000 000 |
| $\frac{1}{2}$ | 0.7291 077 | 0.7290 858 |
| 2 | 0.2987 666 | 0.2987 802 |
| $\frac{9}{2}$ | 0.0832 659 | 0.0832 609 |
| 8 | 0.0152 891 | 0.0152 903 |
| $25\frac{1}{2}$ | 0.0018 126 | 0.0018 124 |
| 18 | 0.0001 368 | 0.0001 368 |
| $49\frac{1}{2}$ | 0.0000 065 | 0.0000 065 |
| 32 | 0.0000 002 | 0.0000 002 |
| Over 32 | 0.0000 000 | 0.0000 000 |
| $\dfrac{E(s^2_{x_R})}{\sigma^2_x}$ | 1.0833 332 | 1.0833 334 |
| $\dfrac{V(s^2_{x_R})}{\sigma^4_x}$ | 2.3430 509 | 2.3430 491 |

\* The values of $u$ shown include every possible value up to 32.

*Position* II: The mean of $x$ in the population coincides with the mid-point of a rounding interval. Thus $\theta = -2\pi i$, where $i$ is some positive or negative integer, or zero.

$$\xi_i = \left(i + \frac{\theta}{2\pi}\right) w$$

$\theta$ may be used to define the location of the rounding lattice. Thus $\theta = 0$ when the mid-point of the 0th interval is at the origin ($x = 0$), and $\theta = \pi$ when the lower boundary of the 0th interval is at the origin. Clearly, the rounding is the same for $\theta = \varphi$ and $\theta = \varphi \pm 2k\pi$, where $k$ is an integer.

Table 4.7 gives to 6 decimal places the exact distributions of $s_{x_R}^2$ for $N = 3$ and $w = \sigma$, corresponding to these two positions of the rounding lattice, when the sampling is from a normal distribution.

TABLE 4.5

EXACT DISTRIBUTIONS OF $s_{x_R}^2 = (x_{R_1} - x_{R_2})^2/2$ FOR THE EXTREME POSITIONS OF THE ROUNDING LATTICE WHEN THE WIDTH OF THE ROUNDING INTERVAL IS ONE-HALF STANDARD DEVIATION AND $x_1$ AND $x_2$ ARE INDEPENDENT OBSERVATIONS FROM A NORMAL POPULATION

$$\left( N = 2, w = \frac{\sigma_x}{2} \right)$$

| $u^*$ | Probability that $\frac{s^2}{\sigma^2} \geq u$ | | $u^*$ | Probability that $\frac{s^2}{\sigma^2} \geq u$ | |
|---|---|---|---|---|---|
| | Position I | Position II | | Position I | Position II |
| 0 | 1.0000 000 | 1.0000 000 | $16\frac{9}{8}$ | 0.0000 110 | 0.0000 110 |
| $\frac{1}{8}$ | 0.8604 037 | 0.8604 037 | $4\frac{9}{2}$ | 0.0000 021 | 0.0000 021 |
| $\frac{1}{2}$ | 0.5977 898 | 0.5977 898 | $22\frac{5}{8}$ | 0.0000 004 | 0.0000 004 |
| $\frac{9}{8}$ | 0.3792 359 | 0.3792 360 | 32 | 0.0000 001 | 0.0000 001 |
| 2 | 0.2183 109 | 0.2183 110 | Over 32 | 0.0000 000 | 0.0000 000 |
| $2\frac{5}{8}$ | 0.1134 753 | 0.1134 754 | $\dfrac{E(s_{x_R}^2)}{\sigma_x^2}$ | 1.0208 312 | 1.0208 314 |
| $\frac{9}{2}$ | 0.0530 515 | 0.0530 515 | | | |
| $4\frac{9}{8}$ | 0.0222 398 | 0.0222 398 | | | |
| 8 | 0.0083 396 | 0.0083 396 | $\dfrac{V(s_{x_R}^2)}{\sigma_x^4}$ | 2.0839 751 | 2.0839 755 |
| $8\frac{1}{8}$ | 0.0027 919 | 0.0027 919 | | | |
| $25\frac{1}{2}$ | 0.0008 331 | 0.0008 331 | | | |
| $12\frac{1}{8}$ | 0.0002 213 | 0.0002 213 | | | |
| 18 | 0.0000 523 | 0.0000 523 | | | |

* The values of $u$ shown include every possible value up to 32.

The probabilities given in these tables were evaluated with the aid of a 15-place table of the probabilities for normal deviates,* and are believed to be correct to within 1 unit in the last place shown. The calculations were carried to 10 decimals or more at all stages and were carefully checked. The individual probabilities obtained, when rounded to the number of places shown, sum to unity within ±1 in the last place shown.

From these tables, and the fact that Positions I and II represent opposite extremes in their effects on $s_{x_R}^2$, the following conclusions may be drawn:

(1) When $N = 2$, the probability that $s_{x_R}^2/\sigma^2$ equals or exceeds $u$ is unaffected to 4 decimal places by the position of the rounding lattice if $w = \sigma$; to 7 decimal places, in general, if $w \leq \sigma/2$.

(2) When $N = 3$, the probability that $s_{x_R}^2/\sigma^2$ equals or exceeds $u$

* Mathematical Tables Project, *Tables of Probability Functions.*

is unaffected by the position of the rounding lattice to 3 decimals if $w = \sigma$.

(3) The effect of the position of the rounding lattice on the distribution of $s_{x_R}^2$ appears to increase with the number of observations involved, that is, with $N$.

TABLE 4.6

EXACT DISTRIBUTIONS OF $s_{x_R}^2 = (x_{R_1} - x_{R_2})^2/2$ FOR THE EXTREME POSITIONS OF THE
ROUNDING LATTICE WHEN THE WIDTH OF THE ROUNDING INTERVAL IS
ONE-QUARTER STANDARD DEVIATION AND $x_1$ AND $x_2$ ARE INDEPENDENT
OBSERVATIONS FROM A NORMAL POPULATION

$$\left(N = 2, \ w = \frac{\sigma_x}{4}\right)$$

| $u^*$ | Probability that $\frac{s^2}{\sigma^2} \geq u$ | | $u^*$ | Probability that $\frac{s^2}{\sigma^2} \geq u$ | |
|---|---|---|---|---|---|
| | Position I | Position II | | Position I | Position II |
| 0 | 1.0000 000 | 1.0000 000 | $8\frac{1}{8}$ | 0.0020 043 | 0.0020 043 |
| $\frac{1}{32}$ | 0.9296 594 | 0.9296 595 | $36\frac{1}{32}$ | 0.0010 902 | 0.0010 902 |
| $\frac{1}{8}$ | 0.7911 480 | 0.7911 481 | $2\frac{5}{2}$ | 0.0005 759 | 0.0005 759 |
| $\frac{9}{32}$ | 0.6589 474 | 0.6589 474 | $44\frac{1}{32}$ | 0.0002 954 | 0.0002 954 |
| $\frac{1}{2}$ | 0.5366 322 | 0.5366 322 | $12\frac{1}{8}$ | 0.0001 471 | 0.0001 471 |
| $25\frac{1}{32}$ | 0.4269 273 | 0.4269 273 | $59\frac{3}{32}$ | 0.0000 711 | 0.0000 711 |
| $\frac{9}{8}$ | 0.3315 445 | 0.3315 445 | 18 | 0.0000 333 | 0 0000 333 |
| $49\frac{3}{32}$ | 0.2511 525 | 0.2511 525 | $62\frac{5}{32}$ | 0.0000 151 | 0.0000 151 |
| 2 | 0.1854 693 | 0.1854 693 | $16\frac{9}{8}$ | 0.0000 066 | 0.0000 066 |
| $8\frac{1}{32}$ | 0.1334 464 | 0.1334 464 | $72\frac{9}{32}$ | 0.0000 028 | 0.0000 028 |
| $2\frac{5}{8}$ | 0.0935 041 | 0.0935 041 | $4\frac{9}{2}$ | 0.0000 011 | 0.0000 011 |
| $12\frac{1}{32}$ | 0.0637 758 | 0.0637 758 | $84\frac{1}{32}$ | 0.0000 004 | 0.0000 004 |
| $\frac{9}{2}$ | 0.0423 269 | 0.0423 269 | $22\frac{5}{8}$ | 0.0000 001 | 0.0000 001 |
| $16\frac{9}{32}$ | 0.0273 253 | 0.0273 253 | Over $22\frac{5}{8}$ | 0.0000 000 | 0.0000 000 |
| $4\frac{9}{8}$ | 0.0171 541 | 0.0171 541 | $\dfrac{E(s_{x_R}^2)}{\sigma_x^2}$ | 1.0052 058 | 1.0052 058 |
| $22\frac{5}{32}$ | 0.0104 691 | 0.0104 691 | | | |
| 8 | 0.0062 099 | 0.0062 099 | $\dfrac{V(s_{x_R}^2)}{\sigma_x^4}$ | 2.0207 869 | 2.0207 869 |
| $28\frac{9}{32}$ | 0.0035 793 | 0.0035 793 | | | |

* The values of $u$ shown include every possible value up to $225/8$.

The mean and variance of $s_{x_R}^2/\sigma_x^2$ for each of the situations considered in Tables 4.4 to 4.7 are given in these tables. These values are brought together in Table 4.8, where they are compared with the corresponding *average* values computed from the formulas of Table 4.2. When $s_{x_R}^2$ is based on $N$ independent (rounded) observations from a normal distribution, then the formula of Table 4.2 for the average variance of $s_{x_R}^2$

becomes

$$\text{average variance of } s_{x_R}^2 = \frac{2\sigma_x^4}{N-1}\left(1 + \frac{a^2}{12}\right)\left(1 - \frac{(N-1)a^4}{240N}\right) \quad (19)$$

where $a = w/\sigma^*$.   Equation (19) gives the average value of the variance of $s_{x_R}^2$ in a series of repeated trials in each of which $N$ independent

TABLE 4.7

EXACT DISTRIBUTIONS OF $s_{x_R}^2 = \sum_{i=1}^{3} (x_{R_i} - \bar{x}_R)^2/2$ FOR THE EXTREME POSITIONS OF

THE ROUNDING LATTICE WHEN THE WIDTH OF THE ROUNDING INTERVAL IS
ONE STANDARD DEVIATION AND $x_1$, $x_2$, AND $x_3$ ARE INDEPENDENT
OBSERVATIONS FROM A NORMAL POPULATION
$(N = 3, w = \sigma_x)$

| $u^*$ | Probability that $\frac{s^2}{\sigma^2} \geq u$ | | $u^*$ | Probability that $\frac{s^2}{\sigma^2} \geq u$ | |
|---|---|---|---|---|---|
| | Position I | Position II | | Position I | Position II |
| 0 | 1.000 000 | 1.000 000 | 9 | 0.000 401 | 0.000 401 |
| ⅓ | 0.915 415 | 0.915 156 | 2⅔ | 0.000 278 | 0.000 278 |
| 1 | 0.541 194 | 0.541 514 | 3⅓ | 0.000 098 | 0.000 098 |
| 4⁄3 | 0.339 213 | 0.338 898 | 12 | 0.000 027 | 0.000 027 |
| 7⁄3 | 0.190 327 | 0.190 245 | 37⁄3 | 0.000 019 | 0.000 019 |
| 3 | 0.071 985 | 0.072 096 | 13 | 0.000 008 | 0.000 008 |
| 4 | 0.040 097 | 0.040 103 | 43⁄3 | 0.000 002 | 0.000 002 |
| 13⁄3 | 0.027 447 | 0.027 411 | 16 | 0.000 001 | 0.000 001 |
| 16⁄3 | 0.008 814 | 0.008 808 | Over 16 | 0.000 000 | 0.000 000 |
| 19⁄3 | 0.005 122 | 0.005 122 | $\dfrac{E(s_{x_R}^2)}{\sigma_x^2}$ | 1.083 330 | 1.083 336 |
| 7 | 0.002 200 | 0.002 206 | | | |
| 25⁄3 | 0.000 629 | 0.000 629 | | | |
| | | | $\dfrac{V(s_{x_R}^2)}{\sigma_x^4}$ | 1.170 804 | 1.170 831 |

\* The values of $u$ shown include every possible value up to 16.

observations are drawn at random from a normal distribution, the position of the rounding lattice being determined at random independently for each trial.   Since the average value of the mean of $s_{x_R}^2$ is unaffected by the nature of the distribution of $x$, its value is given by the formula of Table 4.2 in any case.   From Table 4.8 it is seen that

\* When $N$ is large so that $(N-1)/N \doteq 1$, formula (19) agrees through the term in $a^4$ with the expression given by R. A. Fisher, "Mathematical Foundations," p. 362, for the variance of $\sum_{i=1}^{N} (x_{R_i} - \mu'_{1:x})^2/N$ computed from $N$ independent rounded observations from a normal population.

(1) When $N$ is 2 or 3 and $w \leq \sigma$, the mean values of $s_{x_R}^2$ for Position I and Position II of the lattice agree with each other and with their theoretical average values, to 5 decimal places or more.

(2) When $N$ is 2 or 3 and $w = \sigma$, the variances of $s_{x_R}^2$ for Position I and Position II of the lattice agree with each other and with their theoretical average values to 2 decimal places or more; when $w \leq \sigma/2$, to 3 decimal places or more.

TABLE 4.8

COMPARISON OF ACTUAL VALUES OF THE MEAN AND VARIANCE OF $s_{x_R}^2/\sigma_x^2$ WITH THE VALUES IMPLIED BY THE FORMULAS OF TABLE 4.2 WHEN THE WIDTH OF THE ROUNDING INTERVAL IS ONE STANDARD DEVIATION (AND LESS) AND $s_{x_R}^2$ IS BASED ON TWO OR THREE INDEPENDENT OBSERVATIONS FROM A NORMAL DISTRIBUTION

| Number of observations, $N$ | Width of rounding interval, $\sigma$ units | Lattice position* | Mean value of $s_{x_R}^2/\sigma_x^2$ | | Variance of $s_{x_R}^2/\sigma_x^2$ | |
|---|---|---|---|---|---|---|
| | | | Actual | By formula | Actual | By formula |
| 2 | 1 | I | 1.0833 322 | 1.0833 333 | 2.3430 509 | 2.3423 322 |
| | 1 | II | 1.0833 334 | 1.0833 333 | 2.3430 491 | 2.3423 322 |
| | ½ | I | 1.0208 312 | 1.0208 333 | 2.0839 751 | 2.0839 300 |
| | ½ | II | 1.0208 314 | 1.0208 333 | 2.0839 755 | 2.0839 300 |
| | ¼ | I | 1.0052 058 | 1.0052 083 | 2.0207 869 | 2.0208 711 |
| | ¼ | II | 1.0052 058 | 1.0052 083 | 2.0207 869 | 2.0208 711 |
| 3 | 1 | I | 1.083 330 | 1.083 333 | 1.170 804 | 1.170 351 |
| | 1 | II | 1.083 336 | 1.083 333 | 1.170 831 | 1.170 351 |

* For the exact meanings of Position I and Position II, see Section 5.2.2.

Fisher* has studied the effect of the position of the rounding interval on the mean values of $\bar{x}_R$ and $s_{x_R}^2$ in samples from a normal distribution using an analytical approach based on the periodic character of such effects noted in the first footnote of this section. He has calculated that the bias of $\bar{x}_R$ due to the position of a rounding lattice with intervals one standard deviation wide will be less than one-tenth the standard error of $\bar{x}_R$ for values of $N$ less than $1.379 \times 10^{16}$. For $s_{x_R}^2$ the corresponding value of $N$ is $1.75 \times 10^{14}$; for the variance of $s_{x_R}^2$, it is $1.34 \times 10^{12}$.

From (1) and (2) above and from Fisher's analysis, we conclude that

(3) The effect of the position of the rounding lattice on the mean and variance of $\bar{x}_R$ and $s_{x_R}^2$ in samples from a normal distribution is minute in comparison to the average effects of the rounding.

(4) In small samples (from a normal distribution at least), so far as the approximation of the distributions of $\bar{x}_R$ and $s_{x_R}^2$ by continuous distributions is concerned, the discontinuities of the distributions of $\bar{x}_R$

* FISHER, "Mathematical Foundations," pp. 318–321.

and $s_{x_R}^2$, caused by the rounding, are far more serious than the effects of the position of the rounding lattice.

## 5.3. Combinations of Sample Sizes and Rounding-interval Widths for Which the Actual Distribution of $s_{x_R}^2$ Can Be Approximated Satisfactorily by a Continuous Distribution, When the Sampling Is From a Normal Population

We have not found a rigorous answer to the question of how large the sample size $N$ needs to be for a given rounding-interval width $w$ and how small $w$ needs to be for a given value of $N$ before the $\chi^2$ approximation given in Paragraph (1) of Sec. 4.4 will adequately represent the distribution of $s_{x_R}^2$ for most practical purposes. Since the discontinuities of the distribution of $s_{x_R}^2$ are the principal source of the difficulties encountered in attempting to represent the actual distribution of $s_{x_R}^2$ by a continuous distribution when $N$ is small and $w$ is comparatively large, it is clear that an answer depends on finding the combinations of $N$ and $w$ for which the discontinuities of the distribution of $s_{x_R}^2$ over any given interval are sufficiently numerous and the steps sufficiently small to render them negligible.

TABLE 4.9

PROBABILITY THAT A ZERO VALUE OF $s_{x_R}^2$ WILL BE OBTAINED WHEN $s_{x_R}^2$ IS BASED ON $N$ INDEPENDENT ROUNDED OBSERVATIONS FROM A NORMAL DISTRIBUTION

| Number of observations, $N$ | Width of rounding interval in $\sigma$ units ($a = w/\sigma$) | | | |
|:---:|:---:|:---:|:---:|:---:|
| | 1 | ½ | ¼ | ⅛ |
| 2 | 0.271 | 0.1396 | 0.0703 | 0.0302 |
| 3 | 0.085 | 0.0225 | 0.0057 | 0.0014 |
| 4 | 0.028 | 0.0038 | 0.0005⁻ | 0.0001⁻ |
| 5 | 0.009 | 0.0007 | | |
| 6 | 0.003 | 0.0001⁺ | | |
| 7 | 0.001⁺ | | | |

The combinations of $N$ and $w$ for which the discontinuities of the distribution of $s_{x_R}^2$ are negligible when the sampling is from a normal population could be determined by preparing, for wide ranges of values of $N$ and $w$, tables of the actual distribution of $s_{x_R}^2$ analogous to Tables 4.4 to 4.7. The arithmetical labor of the calculations involved in the preparation of such tables is so enormous, however, that it seems advisable to devise some rule of thumb, which is intuitively sound, as a substitute for rigorous answers until the latter become available. Accordingly, we propose that the probability of $s_{x_R}^2$ equaling zero for given values of $N$ and $w$, when the sampling is from a normal population, be used as a basis for deciding whether a given pair of values of $N$ and $w$ is satisfactory. To

be more explicit, we propose that $P\{s_{x_R}^2 = 0|N,w\} \leq 0.001$, evaluated on the assumption that the sampling is from a normal population, be adopted as a tentative definition of the combinations of values of $N$ and $w$ for which the $\chi^2$ approximation given in (1) of Sec. 4.4 provides a satisfactory approximation to the distribution of $s_{x_R}^2$ for practical purposes.

Values of $P\{s_{x_R}^i = 0|N,w\}$, for sampling from a normal population, are given in Table 4.9 for various combinations of values of $N$ and $w$, the latter being expressed in units of the population standard deviation. It appears from this table that satisfactory combinations, where $a = w/\sigma$, are

$$
\begin{array}{lll}
a \leq \tfrac{1}{200}, & N \geq 2 & \text{(by extrapolation)} \\
a \leq \tfrac{1}{10}, & N \geq 3 & \\
a \leq \tfrac{1}{3}, & N \geq 4 & \quad(20) \\
a \leq \tfrac{1}{2}, & N \geq 5 & \\
a \leq \tfrac{4}{5}, & N \geq 6 & \\
a \leq 1, & N \geq 7 &
\end{array}
$$

The fact that the combinations (20) represent a somewhat arbitrary definition of the circumstances under which (1) and (3) of Sec. 4.4 may be expected to yield satisfactory results in practical work should not be lost sight of. Nor should the combinations (20) be regarded as competing with, or contradicting, the general recommendation, cited in the first paragraph of Sec. 4.1, that $w$ be taken less than $\sigma/3$, or better, less than $\sigma/4$. The object of this general recommendation is to limit the inevitable loss of precision of estimates of parameters caused by rounding, and relates to the case where the choice of a rounding interval is yet to be made. Combinations (20), on the other hand, are not concerned with the loss of precision of estimates as a consequence of coarse rounding but are proposed as a rough indication of the circumstances under which data can be validly analyzed by simple modifications of standard procedures.

### 5.4. Distribution of Student's $t$ and the Variance Ratio When Based on a Small Number of Rounded Observations

Like the distributions of all functions of rounded observations, the distribution of $t_{x_R}$, defined in equation (16) of Sec. 4.5, is discontinuous. However, since the discontinuities of the distribution of $\bar{x}_R$ and the distribution of $s_{x_R}$ jointly affect the distribution of $t_{x_R}$, its discontinuities are more numerous in any given interval than those of the distributions of $\bar{x}_R$ and $s_{x_R}$, and, generally, the steps are smaller than those of the distribution of $s_{x_R}$. On the other hand, whenever $s_{x_R}$ equals zero, this causes $t_{x_R}$ to be either $+\infty$ or $-\infty$, according to the sign of the numerator with the exception of those cases in which $\bar{x}_R$ equals the population mean

$\mu'_{1:x}$, when $t_{x_R}$ can be defined to be equal to $+1$ or $-1$, the sign being determined in a given instance by a random mechanism giving each sign equal probability.* The last-mentioned situation cannot occur unless $\mu'_{1:x}$ is a multiple of $1/N$ and will occur very rarely even when it is a possibility. On the other hand, it can be seen from Table 4.9 that infinite values of $t_{x_R}$ may occur with annoying frequency in very small samples, for example $N \le 4$, when coarse rounding is involved. If, however, the rounding employed with a given $N$ is in accord with the recommendations (20), then the probability of an infinite value of $t_{x_R}$ will be less than 0.001. Furthermore, the rounding so chosen will simultaneously render the steps in the distribution of $t_{x_R}$ negligible, so that for practical purposes $t_{x_R}$ may be regarded as following the standard $t$ distribution, which is strictly valid only for $t_x$.

Similar considerations apply to the distribution of the ratio $F_R$ of two mean squares, $s^2_{1x_R}$ and $s^2_{2x_R}$, which are such that in the absence of the rounding $n_1 s^2_1$ and $n_2 s^2_2$ would be distributed independently as $\chi^2 \sigma^2$ for $n_1$ and $n_2$ degrees of freedom, respectively. In particular, $F_R = s^2_{1x_R}/s^2_{2x_R}$ will equal 0 whenever $s^2_{1x_R}$ is 0; it will equal infinity whenever $s^2_{2x_R}$ is 0; it may be defined to be 1 when both $s^2_{1x_R}$ and $s^2_{2x_R}$ are 0. Four exact distributions of $F_R$ were calculated from the distributions of $s^2_{x_R}$ given in Tables 4.4 to 4.7. The cases considered were $n_1 = n_2 = 1$, $w = \sigma$, $\sigma/2$, $\sigma/4$; $n_1 = n_2 = 2$, $w = \sigma$. As expected, the discontinuities of $F_R$ were more numerous in any given interval than the discontinuities of the distribution of $s^2_{x_R}$ alone, and the steps were smaller. The biggest steps were those corresponding to $F_R = 0$, $F_R = 1$, and $F_R = +\infty$. From the graphs that were sketched it was apparent that if these three steps were reduced to negligible size, either by increasing $n_1$ and $n_2$ or by decreasing $w$, then the distribution of $F_R$ would be represented by the $F$ distribution to an approximation sufficient for most practical purposes. It is believed that if the $N_1 (= n_1 + 1)$, $N_2 (= n_2 + 1)$, and the common rounding interval satisfy the recommendations (20) above, then $F_R$ may be regarded as distributed according to the $F$ distribution for most practical purposes.

If $s^2_{1x_R}$ and $s^2_{2x_R}$ correspond to different roundings $w_1$ and $w_2$, then the distribution of $F_R$ will be biased relative to the distribution of $F$. This bias cannot be eliminated when $\sigma^2$ is unknown, the situation in which $F$ tests are customarily employed.

### 6. STATISTICAL PROPERTIES OF TABULAR AND INTERPOLATED VALUES

It is often stated, or implied, in statistical texts that the error of a tabular entry may be regarded as a random variable distributed with a

---

* The complication of this definition arises because $F = \%$ = 1 seems a sensible definition, and $t = \%$ = 0 also seems sensible, but $t^2 = F$. The complication avoids the implication that $0^2 = 1$.

uniform probability density between minus $\frac{1}{2}$ unit and plus $\frac{1}{2}$ unit in the last place. Clearly the rounding error of a given tabular entry has a definite value, although it may be unknown to a user of the table. Hence the error of a given tabular entry is not a random variable in any meaningful sense of the expression, its distribution being concentrated at a point. Therefore, the statement appearing in statistical texts must be intended to mean that when one is carrying out computations involving the use of a large number of different tabular numbers, the errors of the individual tabular numbers considered collectively will, under favorable circumstances, range from minus $\frac{1}{2}$ unit to plus $\frac{1}{2}$ unit in the last place with approximately uniform density.

The circumstances may sometimes be quite unfavorable. Thus in a given computation involving the use of numbers from only a small section of a table—and in certain other less easily specified cases also— the errors may all tend to have the same sign or to be bunched in some other manner characteristic of the values of the function when expressed to the given number of decimal places for that particular range of the argument. Indeed, the favorability of the circumstances may depend on the units in which the function is tabulated, as the following parallel portions of two logarithm tables illustrate: ＇

| $N$ | $\log_{10} N$ | $\log_e N$ |
|---|---|---|
| 9861 | 3.9939 210 | 9.1963 429 |
| 9862 | - 3.9939 650 | 9.1964 443 |
| 9863 | 3.9940 090 | 9.1965 457 |
| 9864 | 3.9940 531 | 9.1966 470 |
| 9865 | 3.9940 971 | 9.1967 484 |
| 9866 | 3.9941 411 | 9.1968 498 |
| 9867 | 3.9941 851 | 9.1969 511 |
| 9868 | 3.9942 291 | 9.1970 525 |
| 9869 | 3.9942 731 | 9.1971 538 |
| 9870 | 3.9943 172 | 9.1972 551 |
| 9871 | 3.9943 612 | 9.1973 564 |
| 9872 | 3.9944 051 | 9.1974 577 |
| 9873 | 3.9944 491 | 9.1975 590 |
| 9874 | 3.9944 931 | 9.1976 603 |
| 9875 | 3.9945 371 | 9.1977 616 |
| 9876 | 3.9945 811 | 9.1978 629 |
| 9877 | 3.9946 251 | 9.1979 641 |
| 9878 | 3.9946 690 | 9.1980 653 |
| 9879 | 3.9947 130 | 9.1981 666 |

These logarithms are given above correct to 7 decimal places. From an examination of the digits in the seventh place in the case of the values of

$\log_{10} N$, it is apparent that if one is calculating with the aid of logarithms read from a 6-place table of $\log_{10} N$, and the values of $N$ involved are limited to the range shown above, the errors of the logarithms will be bunched around a value of one-tenth unit in the sixth place and quite definitely will not be distributed more or less uniformly from minus ½ unit to plus ½ unit in the sixth place. On the other hand, if the computations are carried out with the aid of a 6-place table of $\log_e N$, then even though the values of $N$ involved are confined to the range shown above, the errors of the logarithms *may* be fairly uniformly distributed between minus ½ unit and plus ½ unit in the sixth place, although there will be a gap in the distribution of the error in the vicinity of two-tenths unit in the last place.

Since there is a definite relationship between the exact values of a function for successive values of the argument, there is likewise a definite relationship between the errors due to rounding in the tabular values given correct to $n$ decimal places. Hence the errors due to rounding to $n$ places are never actually equivalent to independent observed values of a random variable uniformly distributed between minus ½ and plus ½ unit in the last place retained. Nevertheless, if a large number of entries, corresponding to widely dispersed values of the argument, are examined, the figures beyond the $n$th place will often form decimal fractions that appear to be more or less uniformly and independently distributed between 0 and 1 in units of the last figure retained; in some cases, this is true even for relatively small values of $n$. Furthermore, when the function tabulated is irrational or transcendental, the larger the value of $n$, the more nearly do the decimals beyond the $n$th place form decimal fractions that seem to be uniformly and independently distributed, for a given spacing of the argument. There are exceptions, of course. For example, the functions $10^m/9$, $10^m/99$, $10^m/999$, etc. for $m = 0$, $\pm 1$, $\pm 2$, . . . , which have only 0's and 1's in their decimal expansions, do not exhibit either of the aforementioned properties. Similarly, tables of good functions may have some bad patches. However, the fact that the above properties are quite common, especially in the case of decimal expansions of irrational and transcendental numbers, suggests that in many cases it may be useful to regard the errors in tabular numbers due to rounding as uniformly and independently distributed between plus ½ and minus ½ unit in the last place tabulated—provided one exercises reasonable judgment in applying the deductions from this assumption in individual cases.

Let us assume then, with the foregoing reservations, that when the error of a tabular number due to rounding lies between $\pm w/2$, the error $\epsilon$ is distributed with uniform probability density between $\pm w/2$. It follows immediately that

$$\text{Mean error} = \frac{1}{w} \int_{-\frac{w}{2}}^{+\frac{w}{2}} \epsilon \, d\epsilon = 0 \tag{21}$$

$$\sigma_\epsilon^2 = \text{variance of errors} = \frac{1}{w} \int_{-\frac{w}{2}}^{+\frac{w}{2}} \epsilon \, d\epsilon = \frac{w^2}{12} = 0.0833w^2 \tag{22}$$

whence the standard deviation of the errors is

$$\sigma_\epsilon = \frac{w}{\sqrt{12}} = 0.2887w \tag{23}$$

Consequently, if rounding errors are viewed from this statistical viewpoint, equation (21) implies that rounding is unbiased, that is, the average difference between the rounded tabular entry and the true value of the function will be zero. Maintaining the statistical viewpoint, it follows that the sum of the errors of $k$ entries that are "correct" to within $\pm w/2$, *assuming independence of the errors*, will be distributed about a mean of zero with a standard deviation of $w \sqrt{k/12}$; furthermore, the distribution is approximately normal (Gaussian) for $k > 4$.*

As an example of the use of this result, consider a tabulation, supposedly correct to 5 decimals, of the nine terms in the expansion of $(p + q)^8$ for $p = \pi/6$. The sum of the nine values is 1.00003. Does this discrepancy of 0.00003 from the true sum suggest that a mistake has been made in the calculations? The maximum admissible discrepancy is bounded by $\pm 0.000045$, so there is no positive proof that the calculations are in error. The observed discrepancy, measured in units of 0.00001, corresponds approximately to a normal deviate of

$$\frac{(3 - 0)}{\sqrt{9/12}} = 3.46.$$

Since the probability of a normal deviate differing from 0 by more than 2.576 is less than 0.01,† it is advisable to check the calculations (probably starting with the sum).

Fisher and Wishart‡ have used the statistical viewpoint to determine the distribution of the error of an interpolated value when the values of the argument are equally spaced. Let the spacing of the argument be taken as unity, and let 0 to 1 represent the interval containing the value

* For details, see the papers by Baten, Hall, Irwin, and Rietz listed in Sec. 7.

† This may be verified from Table 1.1 of Chap. 1, since the probability is 2(0.005) that a normal deviate will be either above $K_{0.005}$ or below $-K_{0.005}$.

‡ FISHER and WISHART, "On the Distribution of the Error of an Interpolated Value."

of the argument $p$ for which it is desired to obtain an interpolated value; then, if the interpolation is performed without recourse to differences, using Lagrangian interpolation coefficients,* the error of the interpolate can be expressed simply in terms of the tabular entries used to obtain it. It is, in fact, the "interpolate" of the errors:

Error of linear (2-point) interpolate     $= (1 - p)\epsilon_0 + p\epsilon_1$

Error of quadratic (3-point) interpolate

$$= \frac{p(p-1)\epsilon_{-1} + 2(1-p^2)\epsilon_0 + p(p+1)\epsilon_1}{2} \tag{24}$$

Error of $(2m)$-point interpolate        $= \sum_{k=-m+1}^{m} A_k^{2m}(p) \cdot \epsilon_k$

Error of $(2m + 1)$-point interpolate     $= \sum_{k=-m}^{m} A_k^{2m+1}(p) \cdot \epsilon_k$

If the errors of the tabular entries may be regarded as independent random variables with a common variance $\sigma^2$, then the variance of an $r$-point interpolate for argument $p$ is simply

$$V(p,r) = \sum_{k} [A_k^r(p)]^2 \cdot \sigma^2 \tag{25}$$

where the summation is over $k$ from $-m + 1$ to $m$ when $r = 2m$, and over $k$ from $-m$ to $+m$ when $r = 2m + 1$. In particular†

$$V(p,2) = \text{variance of 2-point Lagrangian interpolate} \tag{26}$$
$$= [(1 - p)^2 + p^2]\sigma^2 = [1 - 2p(1 - p)]\sigma^2 \quad 0 \leq p \leq 1$$

and

$$V(p,3) = \text{variance of 3-point Lagrangian interpolate} \tag{27}$$
$$= \frac{[p^2(p-1)^2 + 4(1-p^2)^2 + p^2(p+1)^2]\sigma^2}{4}$$
$$= \frac{[6p^2(p^2 - 1) + 4]\sigma^2}{4} \quad 0 \leq p \leq 1.$$

It can be determined readily by differentiation that the minimum value of $V(p,3)$ corresponds to $p = \pm 1/\sqrt{2}$ and equals

$$\frac{5\sigma^2}{8} = 0.62500\sigma^2$$

---

* For numerical values of these coefficients for 3-point up to 11-point interpolation, see Mathematical Tables Project, *Tables of Lagrangian Interpolation Coefficients;* formulas for these coefficients are given in the introduction to the tables.

† Expression (26) is identical to that given by Fisher and Wishart, p. 914. In place of (27) they give an equivalent expression in terms of a variable $p$ that varies from 0 to 1 when the $p$ of (27) varies from $-\frac{1}{2}$ to $+\frac{1}{2}$.

which exceeds the minimum value of $V(p,2)$, namely $\sigma^2/2 = 0.50000\sigma^2$, which corresponds to $p = \frac{1}{2}$. Also, for $0 \leq p \leq 1$, it is clear that the maxima are $V(0,2) = V(0,3) = V(1,2) = V(1,3) = \sigma^2$. Table 4.10 shows the minima and the average values of $V(p,r)$, $0 \leq p \leq 1$, in the case of $(2m)$-point interpolation $(m = 1, 2, 3, 4)$, and the practical minima* and the average values for $-\frac{1}{2} \leq p \leq \frac{1}{2}$ in the case of $(2m + 1)$-point interpolation $(m = 0, 1, 2, 3)$.

TABLE 4.10

MINIMUM AND AVERAGE VALUES OF THE VARIANCE OF AN $r$-POINT INTERPOLATE FOR AN ARGUMENT BETWEEN 0 AND 1 WHEN $r$ IS EVEN, AND BETWEEN $-\frac{1}{2}$ AND $+\frac{1}{2}$ WHEN $r$ IS ODD, THE VALUES OF THE ARGUMENT BEING SPACED AT UNIT INTERVALS AND THE ERRORS OF THE TABULAR ENTRIES BEING INDEPENDENTLY DISTRIBUTED WITH A COMMON VARIANCE $\sigma^2$

| Interpolate | Minimum Variance* (units of $\sigma^2$) | Average Variance† (units of $\sigma^2$) |
|---|---|---|
| 1-point (an entry) | 1.00000 | 1.00000 |
| 2-point (linear) | 0.50000 | 0.66666 |
| 3-point | 0.71875 | 0.89375 |
| 4-point | 0.64062 | 0.77566 |
| 5-point | 0.74060 | 0.89574 |
| 6-point | 0.70599 | 0.82244 |
| 7-point | 0.76579 | 0.90291 |
| 8-point | 0.74533 | 0.84942 |
| . . . . . . . . | . . . . | . . . . |
| infinite-point | 1.00000 | 1.00000 |

* In the case of $(2m)$-point interpolation, the values shown are actual minima, occurring at $p = \frac{1}{2}$; and in the case of $(2m + 1)$-point interpolation, they are practical minima, occurring at $p = \pm\frac{1}{2}$. For explanation, see text.

† From Fisher and Wishart, pp. 914–915. The average has been taken over the interval $0 \leq p \leq 1$ in the case of $(2m)$-point interpolates; over the interval $-\frac{1}{2} \leq p \leq +\frac{1}{2}$ in the case of $(2m + 1)$-point interpolates.

In summary, if the errors of the tabular entries may be regarded as independent random variables with a common variance $\sigma^2$, if the values of the argument are uniformly spaced at unit intervals, and if $r$-point interpolation for the interpolate corresponding to argument $p$, with $0 < p < 1$ for $r$ even and $-\frac{1}{2} < p < +\frac{1}{2}$ for $r$ odd, is carried on without differences, then

(1) The variance of the random error of the interpolate is always less than the variance of a single tabular entry.

* In the case of $(2m + 1)$-point interpolation for $\frac{1}{2} < p < 1$, a more accurate interpolate can be obtained by taking $p = 1$ as origin, dropping the tabular entry for $k = -m$, adding the entry for $k = m + 1$, and using $1 - p$ as argument. With this convention, $V(p, 2m + 1)$ will have a practical minimum at $p = \pm\frac{1}{2}$, although actual minima occur at $p = \pm p'$ with $|p'| > \frac{1}{2}$.

(2) The variance of the random error of the interpolate is least for linear (2-point) interpolation.

(3) The variance of the random error of a $(2m + 2)$-point interpolate is *greater* than the variance of the random error of a $(2m)$-point interpolate, but *less* than the variance of the random error of a $(2m + 1)$-point interpolate.

(4) The variance of the random error of a $(2m + 3)$-point interpolate is greater than the variance of the random error of a $(2m + 1)$-point interpolate.

From (3) and (4) it is clear that, if $(2m)$-point interpolation does not yield the accuracy desired, but either $(2m + 1)$-point or $(2m + 2)$-point will,* then it may be advisable to employ $(2m + 2)$-point rather than $(2m + 1)$-point interpolation in order to minimize the loss of precision.

It should be noted that these results require merely that the errors of the successive tabular entries be regarded as independently distributed with a common variance—it is not assumed that the errors are due solely to rounding and are distributed with uniform probability density between $-w/2$ and $+w/2$. The results are, of course, applicable to this case, with $\sigma^2 = w^2/12$, as given by (22), when the independence assumption is tenable. Fisher and Wishart point out, however, that in this latter case the distribution of the error of the interpolate is more or less intractable, and does not tend to the normal (Gaussian) distribution ($\beta_2$ tends to 2.2 and not to 3) as $r$ increases. This somewhat limits the usefulness of the standard deviation as a measure of the random error of the interpolate, although the inequalities of Tchebycheff and of Camp and Meidell are still applicable.† Since, by (24), the error of an interpolate is a linear function of the errors of the tabular entries when the interpolation is carried out without recourse to differences (that is, by using Lagrangian interpolation coefficients), the error of the interpolate will be normally distributed when the individual errors are independently and normally distributed with a common variance. Accordingly, they note that a table printed to one or two places more than are correct is capable of yielding a more accurate and more precise interpolate than the same table with the entries cut down, provided that the standard deviation of an entry is given and the errors are approximately normally and independently distributed about a mean of zero with a common variance.

---

* The reader is cautioned against inferring that $(r + 1)$-point, $(r + 2)$-point, and so forth, interpolation always yields greater accuracy than $r$-point interpolation; he is referred to Steffensen, *Interpolation*, pp. 34–53, for a discussion of the number of points to be used in a given instance and for a simple example in which the accuracy becomes worse as $r$ increases.

† See, for instance, SHEWHART, *Economic Control of Quality*, pp. 176–177.

## 7. REFERENCES

ABERNATHY, JOHN R.: "On the Elimination of Systematic Errors due to Grouping," *Annals of Mathematical Statistics*, Vol. 4 (1933), pp. 263–277.

BATEN, W. D.: "Combining Constant Probability Functions," *American Mathematical Monthly*, Vol. 37 (1930), pp. 423–434.

———: "Corrections for the Moments of a Frequency Distribution in Two Variables," *Annals of Mathematical Statistics*, Vol. 2 (1931), pp. 309–319.

CARVER, H. C.: "The Fundamental Nature and Proof of Sheppard's Corrections," *Annals of Mathematical Statistics*, Vol. 7 (1936), pp. 154–163.

CAYLEY, ARTHUR: Letter to J. W. L. Glaisher (1874), reproduced (in part) in *Mathematical Tables and Other Aids to Computation*, Vol. 1 (1943–1945), p. 98.

CRAIG, CECIL C.: "Sheppard's Corrections for a Discrete Variable," *Annals of Mathematical Statistics*, Vol. 7 (1936), pp. 55–61.

ELDERTON, W. PALIN: "Adjustments for the Moments of $J$-Shaped Curves," *Biometrika*, Vol. 25 (1933), p. 179.

FISHER, R. A.: "Applications of 'Student's' Distribution," *Metron*, Vol. 5 (1926) pp. 90–104.

———: "Frequency Distribution of the Values of the Correlation Coefficient in Samples from an Indefinitely Large Population," *Biometrika*, Vol. 10 (1915), pp. 507–521.

———: "On a Distribution Yielding the Error Function of Several Well-known Statistics," *Proceedings of the International Mathematical Congress*, Toronto (1924), Vol. 2, pp. 805–813.

———: "On the Mathematical Foundations of Theoretical Statistics," *Philosophical Transactions of the Royal Society of London*, Vol. 222A (1922), pp. 309–368.

———: *Statistical Methods for Research Workers*, 4th and later ed., Oliver and Boyd, Ltd., Edinburgh and London, 1933–1944.

———, and WISHART, J.: "On the Distribution of the Error of an Interpolated Value, and on the Construction of Tables," *Proceedings of the Cambridge Philosophical Society*, Vol. 23 (1927), pp. 917–921.

GEARY, R. C.: "The Distribution of 'Student's' Ratio for Non-normal Samples," *Supplement to the Journal of the Royal Statistical Society*, Vol. 3 (1936), pp. 178–184.

HALL, PHILIP: "The Distribution of Means of Samples of Size $N$ drawn from a Population in Which the Variate Takes Values between 0 and 1, All Such Values Being Equally Probable," *Biometrika*, Vol. 29 (1927), pp. 240–244.

IRWIN, J. O.: "On the Frequency Distribution of Means of Samples from a Population Having Any Law of Frequency with Finite Moments, with Special Reference to Pearson's Type II," *Biometrika*, Vol. 29 (1927), pp. 226–239.

KENDALL, M. G.: "The Conditions under which Sheppard's Corrections Are Valid," *Journal of the Royal Statistical Society*, Vol. 101 (1938), pp. 592–605.

LANGDON, W. H., and ORE, OYSTEIN: "Semi-Invariants and Sheppard's Corrections," *Annals of Mathematics*, Vol. 31 (1930), pp. 230–232.

Mathematical Tables Project, Work Projects Administration, Federal Works Agency: *Tables of Lagrangian Interpolation Coefficients*, New York, Columbia University Press, 1944, 392 + xxxvi pp.

———: *Tables of Probability Functions, Volume II*, National Bureau of Standards, Washington, D.C., 1942, 344 + xxi pp.

PAIRMAN, ELEANOR, and PEARSON, KARL: "On Corrections for the Moment Coefficients of Limited Range Frequency Distributions When There Are Finite or

Infinite Ordinates and Any Slopes at the Terminals of the Range," *Biometrika*, Vol. 12 (1918–1919), p. 231.

PEARSE, G. E.: "On Corrections for Moment Coefficients of Frequency Distributions When There Are Infinite Ordinates at One or Both Terminals of the Range," *Biometrika*, Vol. 20A (1928), p. 314.

PEARSON, KARL: "Note on Mr. Palin Elderton's Corrections to Moments of J-curves," *Biometrika*, Vol. 25 (1933), p. 180.

RIETZ, H. L. (editor): *Handbook of Mathematical Statistics*, Houghton Mifflin Company, Boston, 1924, 221 + viii pp.

———: "On a Certain Law of Probability of Laplace," *Proceedings of the International Mathematical Congress*, Toronto (1924), Vol. 2, pp. 795–799.

SHEPPARD, W. F.: "On the Calculation of the Most Probable Values of Frequency Constants, for Data Arranged According to Equidistant Divisions of a Scale," *Proceedings of the London Mathematical Society*, Vol. 29 (1898), pp. 353–380.

SHEWHART, W. A.: *Economic Control of Quality of Manufactured Product*, D. Van Nostrand Company, Inc., New York, 1931, 501 + xiv pp.

STEFFENSEN, J. F.: *Interpolation*, The Williams & Wilkins Company, Baltimore, 1927, 248 + ix pp.

STUDENT, "The Probable Error of a Mean," *Biometrika*, Vol. 6 (1908), pp. 1–25; reprinted in *Student's Collected Papers*, ed. by E. S. Pearson and J. Wishart, *Biometrika* Office, University College, London, 1942, 224 + xiv pp.

WOLD, HERMAN: "Sheppard's Correction Formulae in Several Variables," *Skandinavisk Aktuarietidskrift*, Vol. 17 (1934), pp. 248–255.

# CHAPTER 5

# ACCEPTANCE INSPECTION WHEN LOT QUALITY IS MEASURED BY THE RANGE

This chapter was prepared by the editors.

With the exception of Sec. 3, it is based on a memorandum by
M. A. Girshick, Stanford University

# CHAPTER 5

## ACCEPTANCE INSPECTION WHEN LOT QUALITY IS MEASURED BY THE RANGE

### 1. TYPE OF PROBLEM

The range, defined as the difference between the largest and smallest observations in a sample, is familiar in industrial experimentation and quality control as a measure of dispersion in samples. Many applications of the range are to normal or nearly normal populations, for which the variability is better characterized by the standard deviation. The range is used to measure dispersion in samples because of the ease with which it can be determined, and because in small samples it is almost as efficient a measure of variability as the standard deviation.* The present chapter is concerned with an interesting reversal of these roles, where the variability of the lot or process is characterized in terms of the range itself but where the standard deviation is used as the sampling statistic in acceptance inspection to determine whether the range meets specified requirements.

One type of problem in which the range is the parameter of primary interest is that of ensuring the virtual simultaneity of a set of events. This set of events may be thought of as a random sample of $k$ units from an underlying population of such events, and the requirement of simultaneity as a demand that the time difference between the occurrence of the first and last event of such a sample be small. In practice, of course, some samples will meet this requirement and others will not, but we can formulate the requirement statistically as follows: Let the size of the

---

* One statistic is more efficient than another if it varies less from sample to sample for a fixed population, that is, if its sampling variability is less. To get a basis of comparison, it is necessary to convert values of one statistic into implied values of the other. For example, given the range of a sample of $N$ from a normal population, an estimate of the standard deviation can be made with the aid of Tippett's table of the mean range, in units of the population standard deviation, for samples of $N$ from a normal population. Such an estimate will vary more from sample to sample than if the standard deviation had been computed directly by the usual formula. Conversely, given the sample standard deviation, an estimate of the sample range can be formed; such an estimate will vary less from sample to sample than if the range had been computed directly. For samples of 2, however, the range and standard deviation are equally variable; for samples of 3 the range is slightly more variable than the standard deviation, and as the sample size increases the variability of the range increases relative to that of the standard deviation. (For full citations, see Sec. 5.)
Chap. 3, Sec. 4.4, pp. 160–163, discusses the efficiency of statistics.

sample be $k$, and let $t_i$ $(i = 1, 2, \cdots, k)$ be the time of occurrence of the $i$th event of the sample measured from an arbitrary time origin. Denote by $R_k$ the difference between the longest and the shortest of these times in a particular sample.   Then the requirement for an acceptable lot is that in not more than a fraction $\gamma$ of samples of $k$ shall the range $R_k$ exceed a fixed value $U$.   If the time can be measured for each item in the lot, it may be possible to form sets of $k$ in such a way that the range is satisfactory in all or nearly all sets.   On the other hand, if determination of the time alters or destroys the unit, if the time varies from one trial to another for a given unit, or if measurement of the times and systematic grouping is expensive, there are advantages to a procedure such as that suggested, based on random grouping.

A typical problem of ensuring temporal simultaneity arises in connection with mechanisms for which there is a certain lapse of time between activation—for example, closing a switch—and operation.   If such mechanisms are used in assemblies of two or more activated from a single source, it may be important to the functioning of the assembly that these time lapses be nearly equal, in other words, that all the operations occur within a brief interval.   This need can be fulfilled by requiring that each lot of the component mechanisms be such that the range of delay times in random sets of $k$ mechanisms be less than some predetermined constant.   It is not possible, without testing the entire lot, to assure that every set of $k$ formed at random from a lot will have an acceptable range; but it is possible by sampling inspection to obtain any required level of assurance, short of certainty, of detecting lots worse than any specified quality $P$, as measured by the percentage of sets of $k$ in which the range is too large.

Problems of controlling the difference between the largest and smallest measurements in a group of $k$ items are not confined to situations involving temporal variability.   Balance may require that a set of $k$ objects not show too great a range in weight; flatness or straightness may require that a set of $k$ items not vary too much in thickness or length; appearance may depend upon the range of variation within sets of $k$ rather than upon some other type of variability.

### 2. METHOD OF SOLUTION

It is clear that when acceptance inspection is based on samples, there is always a risk that when good lots are offered they will be rejected, or that when bad lots are offered they will be accepted.   The former risk depends on the frequency with which good lots yield bad samples, and the latter on the frequency with which bad lots yield good samples.   These frequencies can be made as low as we please by taking sufficiently large samples and choosing the acceptance criterion properly, but they

cannot be eliminated except by complete inspection of the lot. Following established usage, we shall call the first kind of risk the "producer's risk" and the second kind the "consumer's risk."

For concreteness, let us assume that a lot of acceptable quality is defined as one in which the probability is at least $p_1 = 0.99$ that the range of 14 items selected at random will be less than 0.3, measured in whatever units are appropriate. Let us suppose further that the producer's risk is to be a small quantity, say $\alpha = 0.01$; that is to say, if lots of quality $p_1$ are offered we wish to reject them, on the average, no oftener than 1 time in 100. Suppose that a lot is unacceptable if the probability is $p_2 = 0.85$, or less, that the range of 14 items selected at random will be less than 0.3; and that the consumer's risk—the probability of accepting a lot of quality $p_2$ or worse if offered—is to be at most $\beta = 0.01$. Finally, let us assume that the distribution of the quality characteristic is adequately represented by a normal distribution.*

For a normal distribution the probability that the range of $k$ independent observations is less than some preassigned number $U$ depends only on the standard deviation $\sigma$. Thus, the criteria of acceptable and unacceptable lots, which were expressed above in terms of the proportion of the ranges of sets of 14 that are less than 0.3, can be expressed in terms of the lot standard deviation $\sigma$. It follows that the acceptance-inspection plan can be formulated in terms of the sample standard deviation, or, as proves to be more convenient in practice, in terms of the sum and the sum of squares of the observations. This reformulation of the definition of an acceptable lot permits a great gain in the efficiency of the inspection plan; for in using the sum of squared deviations as a measure of acceptability we make full use of the information provided by inspection of the sample, rather than of just that portion of the information given by selecting the extreme individuals in each group of 14 observations.

Let $\sigma_1$ be the value of the lot standard deviation that defines good quality. Then $\sigma_1$ depends on the group size $k$, on the acceptable value $U$ of the range in groups of $k$, and on the proportion of ranges that can be allowed to exceed $U$. These three quantities are related to $\sigma_1$ through the cumulative distribution of the standardized range, which gives the probability that the range $R_k$ of $k$ items from a normal distribution will exceed any specified multiple $W_k$ of the population standard deviation $\sigma$. Tables of the cumulative distribution of the range have been published by Pearson.† In Fig. 5.1, which is based on Pearson's

---

* If the assumption of normality is not justified, the inspection can be carried out on an attributes basis, by classifying each group of 14 in the sample as satisfactory or not.

† Pearson's table gives cumulative probabilities to 4 decimal places for values of $k$

tables, the vertical axis shows the probability that $R_k/\sigma$ will exceed the values of $W$ shown on the horizontal axis, and the various curves correspond with various group sizes $k$. For our purposes, $R_k$ is given, $W$ is obtained from Pearson's table or from Fig. 5.1, and $\sigma$ is the quotient

$$\sigma = \frac{R_k}{W_k} \tag{1}$$
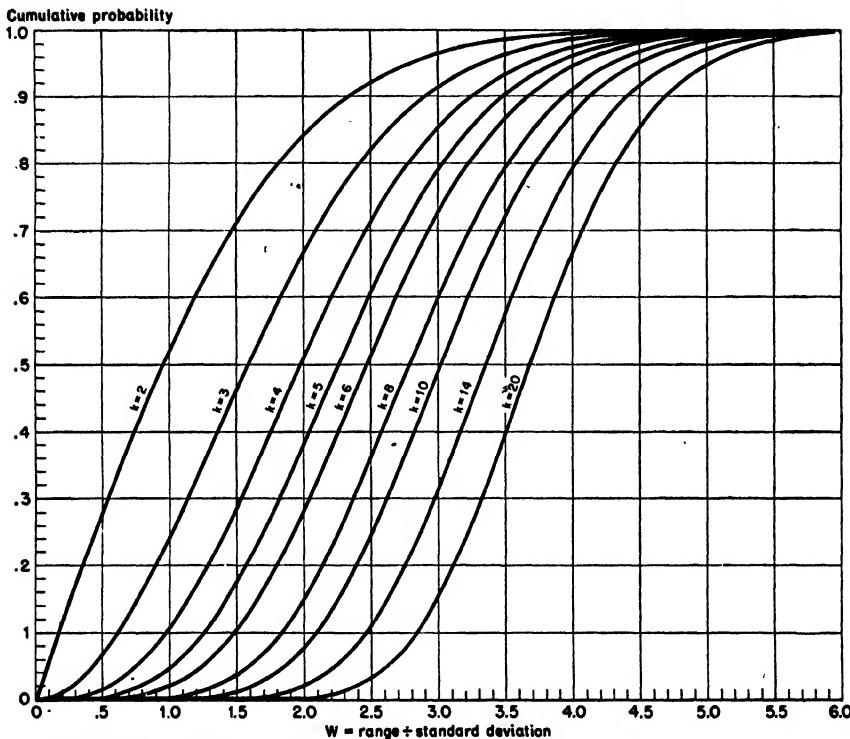


Data for these curves are from E. S. Pearson, *Biometrika*, 1942.-

Fig. 5.1.—Cumulative probability distribution of the range in samples of $k$ from a normal population.

For the illustration cited, we find the value of $W_{14}$ corresponding with probabilities of 0.99 and 0.85 on the curve for $k = 14$. These are 5.40 and 4.20, respectively; on substituting them and $R_k = 0.3$ into (1), we find $\frac{1}{18}$ and $\frac{1}{14}$ as the values of $\sigma_1$ and $\sigma_2$.

Thus the problem becomes that of devising an inspection procedure for the standard deviation such that the operating-characteristic curve passes through the two points

(in his notation $n$) from 2 to 20 by steps of 1 and for values of $W$ from 0 to 7.25 by steps of 0.01.

$$\sigma_1 = 0.05556 \qquad L_{\sigma_1} = 0.99 \tag{2}$$
$$\sigma_2 = 0.07143 \qquad L_{\sigma_2} = 0.01 \tag{3}$$

Such an inspection procedure may be carried out by either single or sequential sampling. The solutions are indicated only briefly here.[*]

## 3. SINGLE SAMPLING

To determine the required sample size, we compute[†]

$$\rho = \left(\frac{\sigma_2}{\sigma_1}\right)^2 = 1.6531 \tag{4}$$

Then in a table of $\chi^2$ we find the number of degrees of freedom $n$ for which

$$\frac{\chi_\alpha^2}{\chi_{1-\beta}^2} = \rho \tag{5}$$

where $\chi_\epsilon^2$ denotes the value of $\chi^2$ exceeded with probability $\epsilon$.

In the present example, since $\alpha = \beta = 0.01$, we may refer to Table 8.2 and seek 1.6531 in the column for $\beta = 0.01$; however, the table shows only that $n > 120$. We therefore use the large sample approximation[‡]

$$n = \frac{1}{2}\left(\frac{K_\alpha \sigma_1 + K_\beta \sigma_2}{\sigma_2 - \sigma_1}\right)^2 \tag{6}$$
$$= 173$$

where $K_\alpha$ and $K_\beta$ represent normal deviates exceeded with probabilities $\alpha$ and $\beta$, respectively, which may be found from Table 1.1 in Chap. 1. Thus 173 is the number of degrees of freedom required, and since the mean is unknown the required sample size is

$$N = n + 1 \tag{7}$$
$$= 174$$

From the sample of 174, we calculate

$$S = \sum_{i=1}^{N} x_i^2 - \frac{(\Sigma x_i)^2}{N} \tag{8}$$

where $x_i$ ($i = 1, 2, \cdots, N$) represents the value of the $i$th measurement in the sample. The lot is rejected if $S > S_0$, where

$$S_0 = \sigma_1^2 \chi_\alpha^2(n) \tag{9}$$

[*] For a discussion of tests of the standard deviation see, on single sampling, Chap. 8, and, on sequential sampling, Chap. 17 below and Statistical Research Group, *Sequential Analysis*, Sec. 6.

[†] See Chap. 8, Sec. 2.4, especially Example III, pp. 274–278.

[‡] See Chap. 8, Sec. 2.4, first footnote on p. 278.

in which $\chi_{\alpha}^2(n)$ represents the value of $\chi^2$ for $n(= N - 1)$ degrees of freedom exceeded with probability $\alpha$; otherwise the lot is accepted. In the present example, $n$ is outside the range of the $\chi^2$ tables, so we use the approximation (due to R. A. Fisher)

$$\chi_{\alpha}^2(n) = \frac{(K_{\alpha} + \sqrt{2n - 1})^2}{2} \tag{10}$$

$$= 218.416$$

The OC curve is easily obtained from the fact that the probability $L_{\sigma}$ of accepting a lot of quality $\sigma$ when offered is simply the probability that a value of $\chi^2$ based on $n$ degrees of freedom will be less than $S_0/\sigma^2$. When $\alpha = 0.01$ or $0.05$ and $n$ does not exceed 120, Table 8.1 or 8.2 provides 13 points on the OC curve; values of $\sigma^2/\sigma_1^2$ are shown for various values of $L_{\sigma}$. When $n$ is beyond the range of the table, the OC curve can be found from

$$\sigma = \frac{(K_{\alpha} + \sqrt{2n})\sigma_1}{\sqrt{2n} - K_{L_{\sigma}}} \tag{11}$$

### 4. SEQUENTIAL SAMPLING

A sequential-inspection plan based on standard deviations is defined by two parallel straight lines given by the formulas

$$Z_n = h_2 + sn \qquad \text{(rejection line)} \qquad (12)$$
$$Z_n = -h_1 + sn \qquad \text{(acceptance line)} \qquad (13)$$

where $h_1$, $h_2$, and $s$ are positive, $n$ represents one less than the number of observations taken sequentially, and

$$Z_n = \sum_{i=1}^{n+1} x_i^2 - \frac{(\Sigma x_i)^2}{n + 1}$$

is the sum of squares of deviations of the $n + 1$ observations from their mean.

The sequential procedure is as follows: Draw the two parallel lines given by equations (12) and (13) with $Z_n$ on the vertical scale and $n$ on the horizontal scale. Start taking observations, beginning with two observations and thereafter adding one observation at a time. After each observation compute $Z_n$ from all observations and plot its value against the corresponding value of $n$ (which is one less than the total number of observations thus far taken). If at any stage the plotted point falls on or above the upper line, given by equation (12), reject the lot; if it falls on or below the lower line, given by equation (13), accept the lot. If, however, the plotted point falls between the two parallel lines, take another observation. Continue this process until a decision of acceptance or rejection is reached.

The two intercepts $-h_1$ and $h_2$ and the common slope $s$, which determine the two parallel lines, are given by

$$h_1 = \frac{2 \log_e \dfrac{1 - \alpha}{\beta}}{\dfrac{1}{\sigma_1^2} - \dfrac{1}{\sigma_2^2}} \tag{14}$$

$$= 0.071799$$

$$h_2 = \frac{2 \log_e \dfrac{1 - \beta}{\alpha}}{\dfrac{1}{\sigma_1^2} - \dfrac{1}{\sigma_2^2}} \tag{15}$$

$$= 0.71799$$

and

$$s = \frac{\log_e \dfrac{\sigma_2^2}{\sigma_1^2}}{\dfrac{1}{\sigma_1^2} - \dfrac{1}{\sigma_2^2}} \tag{16}$$

$$= 0.0039268$$

The OC curve for the sequential plan is computed from the following formulas:

$$\sigma = \sqrt{\frac{1 - e^{-2ts}}{2t}} \qquad L_\sigma = \frac{e^{t(h_1 + h_2)} - e^{th_1}}{e^{t(h_1 + h_2)} - 1} \tag{17}$$

It will be observed that $L_\sigma$, the ordinate of the OC curve, is not given directly as a function of $\sigma$, but parametrically through a third variable $t$ in terms of which $\sigma$ is also defined. The OC curve for this example is shown in Fig. 5.2.

The average sample number (ASN) curve shows the relationship between the standard deviation in the submitted lot and the average number of items that will be inspected before a decision is reached. The ordinate $N_\sigma$ of the ASN curve, corresponding to the abscissa $\sigma$, is given by

$$\bar{N}_\sigma = \frac{L_\sigma(h_1 + h_2) - h_2}{s - \sigma^2} + 1 \tag{18}$$

where $L_\sigma$ and $\sigma$ are given by equations (17). Formula (18) can be evaluated directly for every value of $\sigma$ except $\sigma = \sqrt{s}$, for which it can be shown that

$$\bar{N}_{\sqrt{s}} = \frac{h_1 h_2}{2s^2} + 1 \tag{19}$$

This will ordinarily be close to the maximum average amount of inspection required to reach a decision. The ASN curve for the present example is included in Fig. 5.2.
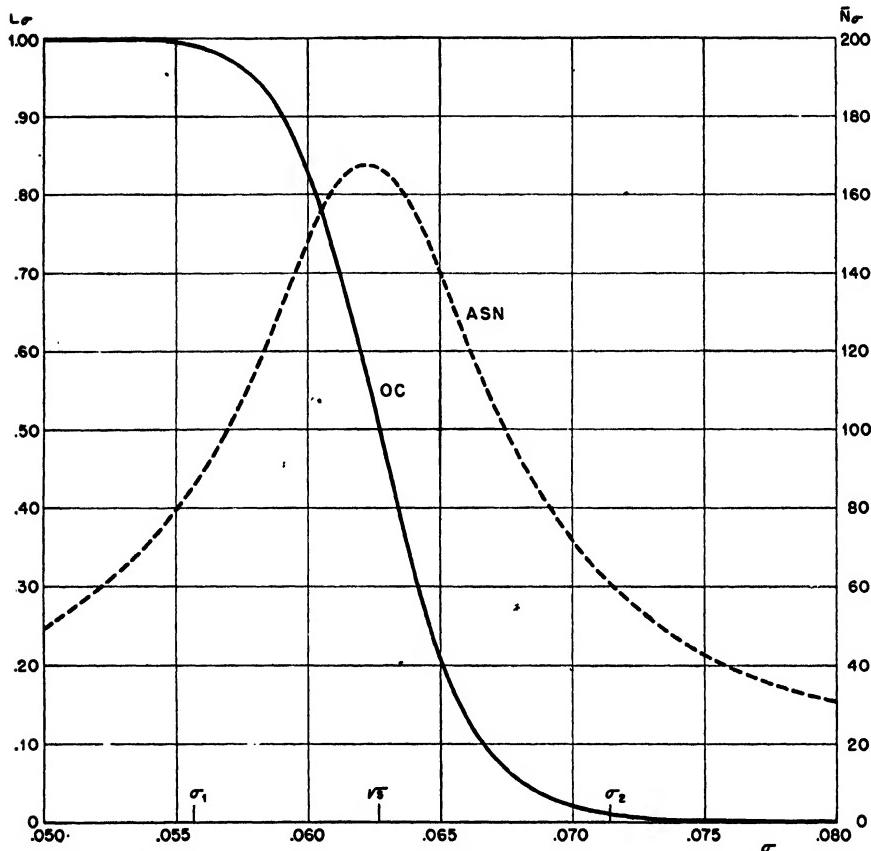
FIG. 5.2.—OC and ASN curves for sequential-sampling plan based on standard deviation; $\sigma_1 = 0.05556$, $\sigma_2 = 0.07143$, $\alpha = 0.01$, $\beta = 0.01$; $h_1 = 0.07180$, $h_2 = 0.07180$, $s = 0.003927$.

## 5. REFERENCES

FISHER, RONALD A., and YATES, FRANK: *Statistical Tables for Biological, Agricultural, and Medical Research*, 2d ed., Oliver and Boyd, Ltd., London and Edinburgh, 1943, 98 + viii pp.

PEARSON, E. S.: "The Probability Integral of the Range in Samples of $n$ Observations from a Normal Population," *Biometrika*, Vol. 32 (1942), pp. 301–308.

Statistical Research Group, Columbia University: *Sampling Inspection*, McGraw-Hill Book Company, Inc., New York, 1948.

Statistical Research Group, Columbia University: *Sequential Analysis of Statistical Data: Applications*, rev. ed., Columbia University Press, New York, 1945.

THOMPSON, CATHERINE M.: "Table of Percentage Points of the $\chi^2$ Distribution," *Biometrika*, Vol. 32 (1941), pp. 187–191.

TIPPETT, L. H. C.: "The Distribution of the Extreme Individuals and of the Range in Samples from a Normal Population," *Biometrika*, Vol. 17 (1925), pp. 364–387; *Tables for Statisticians and Biometricians*, Part II, Table XXII, pp. 165–166.

# CHAPTER 6

# VARIABILITY OF AMOUNT OF INSPECTION
## FOR DOUBLE, MULTIPLE, AND SEQUENTIAL SAMPLING

### FIGURES

### TABLES

235

This chapter was prepared by the editors
from more extensive material prepared by
Myra Levine, Socony Vacuum Oil Company

# CHAPTER 6

## VARIABILITY OF AMOUNT OF INSPECTION FOR DOUBLE, MULTIPLE, AND SEQUENTIAL SAMPLING

### 1. SINGLE, DOUBLE, MULTIPLE, AND SEQUENTIAL SAMPLING

Sampling inspection is said to be based on attributes if inspection of each item results in its classification as defective or nondefective, and disposition of the lot or process from which the sample comes depends upon the number of defectives found in the sample. One or more samples are selected at random from the lot; if the total number of defectives found in the first $k$ samples does not exceed an acceptance number $A_k$, the lot is accepted; if it equals or exceeds a rejection number $R_k$ $(R_k > A_k)$, the lot is rejected; if it falls between $A_k$ and $R_k$, another sample is inspected.

A single-sampling plan is one for which

$$R_1 = A_1 + 1 \tag{1}$$

A double-sampling plan is one for which

$$R_1 > A_1 + 1$$
$$R_2 = A_2 + 1 \tag{2}$$

A multiple-sampling plan is one for which

$$R_k > A_k + 1 \qquad \text{for } k \leq K \tag{3}$$
$$R_k = A_k + 1 \qquad \text{for some } K > 2$$

In a sequential-sampling plan, the successive samples are all of size 1 and

$$R_k - A_k = C + \delta_k \tag{4}$$

where $C$ is some constant and each $\delta_k$ is 0 or 1.[*]

---

[*] For an exposition of single, double, and multiple sampling, see Statistical Research Group, *Sampling Inspection;* what is here called "multiple sampling" is there called "sequential sampling," and what is here called "sequential sampling" is there called "item-by-item sequential sampling." For an exposition of sequential sampling see Chap. 17 of this volume, or Statistical Research Group, *Sequential Analysis.* (For full citations, see Sec. 5.)

Equation (4) is a necessary condition for a sequential-sampling plan, but it is not quite a sufficient condition unless further conditions are imposed on the $\delta_k$.

## 2. VARIABILITY OF AMOUNT OF INSPECTION

One of several factors determining the relative merits of single, double, multiple, and sequential sampling is the variability of the amount of inspection for plans giving equivalent operating-characteristic (OC) curves. Variability of the amount of inspection causes fluctuations in the inspection load and thus may create administrative problems in an inspection program.

Variability may be interpreted in two ways. First, it may mean variation in the average number of items inspected per lot as the quality of submitted lots varies; in other words, the amplitude of the rise and fall of the average sample number (ASN) curve. Second, it may mean variation in the number of items inspected for lots of a given quality.

Clearly, single sampling will show less variability, in either sense, than double, multiple, or sequential sampling. In fact, with single sampling there will be no variability of either kind unless inspection is curtailed, that is, terminated as soon as the rejection number is exceeded or the number of items remaining uninspected is insufficient to cause rejection even if all are defective.

No general relations among the variabilities of double, multiple, and sequential plans are known. The purpose of this chapter is to present two illustrative comparisons. These two comparisons suggest by their differences that no simple conclusions about the relative variability of the three types of sampling are valid. In the illustrations it is assumed that the first sample of the double and multiple plans will be inspected completely (this being usual* to facilitate estimation of the process average), that inspection of later samples will be curtailed as soon as the rejection number is reached, but that curtailing will not be applied to accepted lots.

## 3. ILLUSTRATIONS

*Example* I. Table 6.1 presents a set of double, multiple, and sequential plans whose OC curves are nearly the same.† Figure 6.1 shows their average sample numbers $\bar{N}$ as functions of lot quality, and Fig. 6.2 shows their standard deviations $\sigma(N)$ of amount of inspection for lots of a given quality as functions of lot quality.

*Example* II. Table 6.2 presents a second set of plans with matching OC curves;‡ Fig. 6.3 shows $\bar{N}$ as a function of lot quality, and Fig. 6.4 shows $\sigma(N)$ as a function of lot quality.

---

* It is not, however, necessary. See GIRSHICK, MOSTELLER, and SAVAGE, "Unbiased Estimates for Certain Binomial Sampling Problems with Applications."

† The OC curves of the double and multiple plans are shown in Table 4 of *Sampling Inspection* for sample size letter $H$, acceptable quality level (AQL) class 0.32–0.65 percent defective (p. 288).

‡ The OC curves of the double and multiple plans are shown in Table 4 of *Sampling Inspection* for sample size letter $K$, AQL class 3.2–4.4 percent defective (p. 330).

## TABLE 6.1
### DOUBLE, MULTIPLE, AND SEQUENTIAL PLANS OF EXAMPLE I

DOUBLE SAMPLING

| Sample | Sample size | Combined samples | | |
|---|---|---|---|---|
| | | Size | Acceptance number | Rejection number |
| First..................... | 50 | 50 | 0 | 3 |
| Second................... | 100 | 150 | 2 | 3 |

MULTIPLE SAMPLING

| Sample | Sample size | Combined samples | | |
|---|---|---|---|---|
| | | Size | Acceptance number | Rejection number |
| First........................ | 20 | 20 | * | 2 |
| Second...................... | 20 | 40 | * | 2 |
| Third....................... | 20 | 60 | 0 | 2 |
| Fourth...................... | 20 | 80 | 1 | 3 |
| Fifth........................ | 20 | 100 | 1 | 3 |
| Sixth....................... | 20 | 120 | 1 | 3 |
| Seventh..................... | 20 | 140 | 2 | 3 |

* Acceptance not permitted until three samples have been inspected.

SEQUENTIAL SAMPLING

$$\alpha = 0.05 \qquad \beta = 0.10 \qquad p_1 = 0.005 \qquad p_2 = 0.05$$
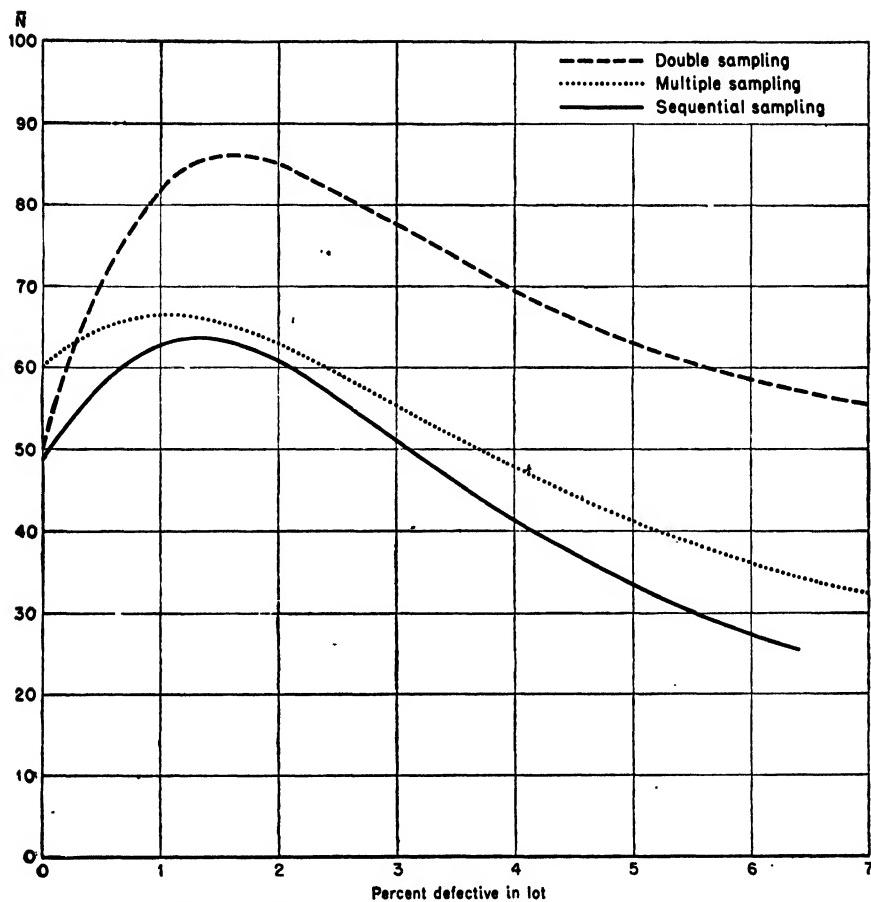
FIG. 6.1.—ASN curves for double, multiple, and sequential plans of Table 6.1.
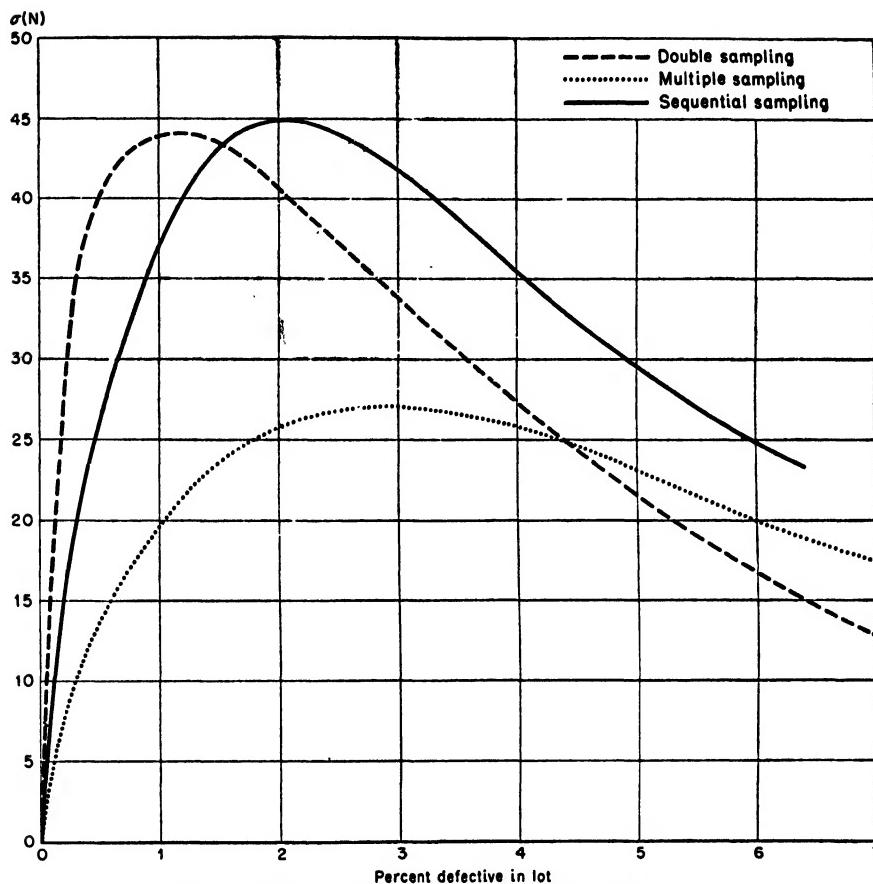
FIG. 6.2.—$\sigma(N)$ curves for double, multiple, and sequential plans of Table 6.1.
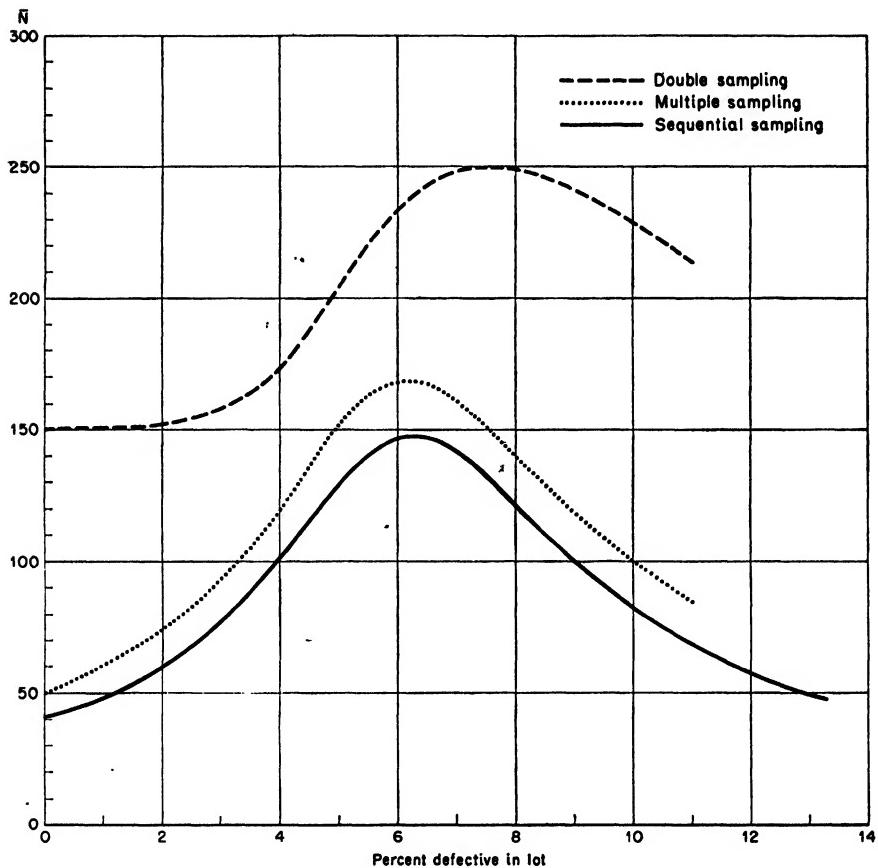
FIG. 6.3.—ASN curves for double, multiple, and sequential plans of Table 6.2.
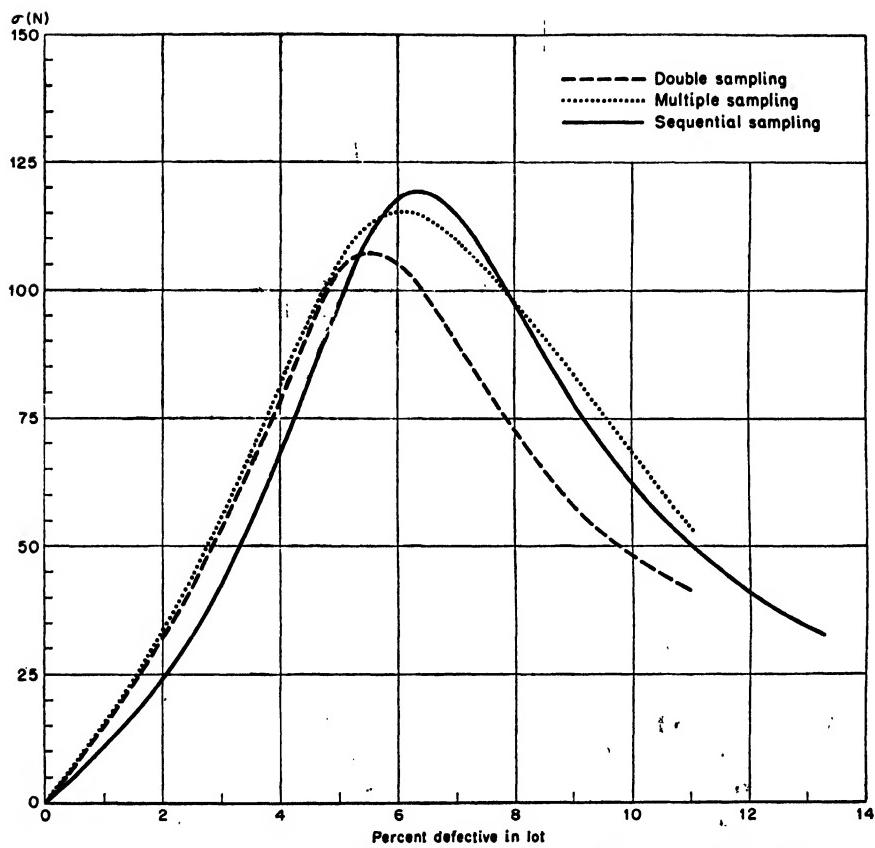
FIG. 6.4.—$\sigma(N)$ curves for double, multiple, and sequential plans of Table 6.2.

TABLE 6.2
DOUBLE, MULTIPLE, AND SEQUENTIAL PLANS OF EXAMPLE II

DOUBLE SAMPLING

| Sample | Sample size | Combined samples | | |
|---|---|---|---|---|
| | | Size | Acceptance number | Rejection number |
| First........................ | 150 | 150 | 9 | 24 |
| Second....................... | 300 | 450 | 23 | 24 |

MULTIPLE SAMPLING

| Sample | Sample size | Combined samples | | |
|---|---|---|---|---|
| | | Size | Acceptance number | Rejection number |
| First........................ | 50 | 50 | 1 | 6 |
| Second....................... | 50 | 100 | 3 | 9 |
| Third........................ | 50 | 150 | 7 | 13 |
| Fourth....................... | 50 | 200 | 10 | 16 |
| Fifth........................ | 50 | 250 | 13 | 19 |
| Sixth........................ | 50 | 300 | 16 | 22 |
| Seventh...................... | 50 | 350 | 19 | 25 |
| Eighth................. ..... | 50 | 400 | 24 | 25 |

SEQUENTIAL SAMPLING

$\alpha = 0.05$    $\beta = 0.10$    $p_1 = 0.0410$    $p_2 = 0.0920$

## 4. COMPUTATIONAL METHODS

For double and multiple sampling, $\bar{N}$ and $\sigma(N)$ may be calculated by an extension of methods described in *Sampling Inspection*, Chap. 17. For sequential sampling, the computations for $\bar{N} = E(N)$ are based on the formula derived in *Sequential Analysis*, Sec. B.0921, and those for $\sigma(N)$ on the relation

$$\sigma^2(N) = E(N^2) - [E(N)]^2 \tag{5}$$

where the formula for $E(N^2)$ may be obtained by an extension of the method used in Sec. B.0921 of *Sequential Analysis* for obtaining $E(N)$.

## 5. REFERENCES

GIRSHICK, M. A., MOSTELLER, F., and SAVAGE, L. J.: "Unbiased Estimates for Certain Binomial Sampling Problems with Applications," *Annals of Mathematical Statistics*, Vol. 17 (1946), pp. 13–23.

Statistical Research Group, Columbia University: *Sampling Inspection*, McGraw-Hill Book Company, Inc., New York, 1948.

———: *Sequential Analysis of Statistical Data: Applications*, rev. ed., Columbia University Press, New York, 1945.

# *Part* II

# PLANNING EXPERIMENTS

# CHAPTER 7

# PLANNING AND ANALYZING EXPERIMENTS FOR COMPARING TWO PERCENTAGES

## FIGURES

This chapter was prepared by
Edward Paulson, University of Washington
W. Allen Wallis, University of Chicago

# CHAPTER 7

# PLANNING AND ANALYZING EXPERIMENTS FOR COMPARING TWO PERCENTAGES

## 1. TYPE OF PROBLEM

The type of problem dealt with in this chapter may be indicated by a simple example. Suppose that trials of two fire-control devices have resulted in the following data:

TABLE 7.1

|  | Hits | Misses | Total |
|---|---|---|---|
| Old........................... | 3 | 197 | 200 |
| New........................... | 4 | 196 | 200 |
| Total........................... | 7 | 393 | 400 |

The question is whether the new method, which shows 33 percent more hits than the old, can be expected to yield more hits in the long run; or whether its superiority in this experiment ought to be regarded as mere chance, in the sense that two series of trials with the same method would often differ as much.

We may look at the question this way: From 400 rounds, 7 hits were secured. If the two methods are really equally good in the long run, these 7 hits should be divided between the two series in proportion to the total numbers in the series, except for chance variations. In the present case we should expect the 7 hits to be divided equally between the two series; and similarly the 393 misses. As it happens, the division in Table 7.1 is as nearly equal as it possibly could be, since the entries are necessarily integers. It is therefore immediately obvious in this particular instance that the probability is one-half that, if the two methods were really equally good in the long run, a single experiment would show for the new method as much (or more) apparent superiority as is shown in Table 7.1. (If this is not obvious, it can be verified by the method described below for use with cases where the answer is not apparent.)

Data such as those of Table 7.1 represent a double dichotomy. The observations are dichotomized according to one criterion—that is, they are divided into two mutually exclusive and exhaustive classes—and then each class is dichotomized again according to a second criterion.

The result of doubly dichotomizing the data is, of course, the same whichever criterion is applied first; that is, we may regard the data as classified first by method (old or new) and then by effect (hit or miss), or we may regard them as classified first by effect and then by method. The marginal totals of Table 7.1 represent two independent dichotomies of the data; they do not constitute a double dichotomy. The problem may be put in either of two ways: (1) Is the division between effects the same for both methods? (2) Is the division between methods the same for both effects? These questions are logically identical, though for a particular problem one of them usually seems more natural than the other. The problem is thus to test the homogeneity of a pair of dichotomies.

## 2. METHOD OF ANALYSIS*

### 2.1. Exact Method

In order to illustrate the general procedure for testing the homogeneity of a double dichotomy such as that of Table 7.1, consider the data of Table 7.2, where the answer is not immediately apparent.

TABLE 7.2

|  | Successes | Failures | Total |
|---|---|---|---|
| Standard...................................... | 2 | 198 | 200 |
| Experimental................................. | 5 | 195 | 200 |
| Total...................................... | 7 | 393 | 400 |

For convenience and generality the two methods compared are here designated "standard" and "experimental" and the two possible outcomes of a trial with either method as "success" and "failure." In practice, of course, the two methods may be equally experimental or equally standard, and the two possible outcomes may be two types of success or two types of failure or merely two different outcomes not properly regarded as either success or failure.

It might appear, in view of the remarks in Sec. 1, that we could treat the 2:5 division of successes like a 2:5 division of heads and tails on 7 tosses of a coin, using the binomial distribution to find the probability of as many as 5 heads on 7 tosses if the coin were really unbiased. Table 7.2, unlike Table 7.1, however, gives different answers for the successes and failures columns according to this treatment; so some other method of analysis is called for. (In this particular instance, however, it happens that analysis of the seven successes according to a binomial distribution gives an excellent approximation to the correct probability, because both proportions of successes are small; see below, Sec. 2.2, p. 255.)

We proceed as follows. Let $P$ designate the true long-run probability

* See FISHER, *Statistical Methods*, Secs. 21, 21.01, and 21.02. (For full citations, see Sec. 6.)

of a success, that is, the probability of a success in the population from which the sample comes. $P$, of course, is unknown, for we should have to carry out an infinitely long series of trials to measure it. But we make our calculations on the assumption that $P$ is the same in both series, and, as we shall see, it cancels out. If, on the assumption that $P$ is really the same in the long run for both series, we find that a difference as great as that shown in the data would be unlikely, then the fact that this unlikely difference has actually been observed experimentally leads us to conclude that the assumption was wrong—to conclude, in other words, that the long-run probabilities are not equal.

The probability that 200 trials will show exactly 2 successes is, by the binomial distribution,

$$\frac{200!}{2!198!} P^2 (1 - P)^{198} \tag{1}$$

Similarly, the probability of exactly 5 successes is

$$\frac{200!}{5!195!} P^5 (1 - P)^{195} \tag{2}$$

So the probability that the two series of 200 trials will show exactly 2 successes for the first series and exactly 5 for the second is the product of (1) and (2)

$$\frac{200!200!}{2!5!195!198!} P^7 (1 - P)^{393} \tag{3}$$

This is an absolute probability, whereas we require the conditional probability that if exactly 7 successes and 393 failures occur in the two series they will be divided between the series as shown in Table 7.2; hence (3) must be divided by the probability of exactly 7 successes in 400 trials.

$$\frac{400!}{7!393!} P^7 (1 - P)^{393} \tag{4}$$

The probability that if we get 7 successes among two series of 200 each, they will be distributed 2 to the standard method and 5 to the experimental, is the ratio of (3) to (4), or

$$\frac{7!393!200!200!}{400!2!5!195!198!} = 0.1636 \tag{5}$$

It should be noted that expression (5) would be the same if expressions (1), (2), and (4) had been formed from the two columns instead of from the two rows of Table 7.2.

In general, expression (5) is composed as follows: The numerator is the product of the factorials of the four marginal totals; the denominator is the product of the factorials of the four cell entries and the factorial

of their sum.    Thus, if we give Tables 7.1 and 7.2 the general representation shown in Table 7.3,

TABLE 7.3

|  | Successes | Failures | Total |
|---|---|---|---|
| Standard............................... | $a$ | $b$ | $a + b$ |
| Experimental........................... | $c$ | $d$ | $c + d$ |
| Total................................... | $a + c$ | $b + d$ | $a + b + c + d$ |

expression (5) may be written

$$\frac{(a + b)!(c + d)!(a + c)!(b + d)!}{a!b!c!d!(a + b + c + d)!} \tag{6}$$

Expression (5) gives the probability that if in fact the long-run probabilities were equal, an experiment involving 200 trials of each method and resulting in a total of 7 successes would show exactly the superiority for the experimental method that is shown by Table 7.2.    For the final probability to use in judging whether there is a long-run difference, we must add to (5) the probabilities of superiorities greater than that of Table 7.2.    Tables 7.4 and 7.5 show possible results that would have indicated even greater superiority for the experimental method than does Table 7.2.

TABLE 7.4

|  | Successes | Failures | Total |
|---|---|---|---|
| Standard....................................... | 1 | 199 | 200 |
| Experimental.................................... | 6 | 194 | 200 |
| Total........................................... | 7 | 393 | 400 |

Expression (6) gives the probability for Table 7.4 as

$$\frac{7!393!200!200!}{400!1!6!194!199!} = 0.0534 \tag{7}$$

TABLE 7.5

|  | Successes | Failures | Total |
|---|---|---|---|
| Standard....................................... | 0 | 200 | 200 |
| Experimental.................................... | 7 | 193 | 200 |
| Total........................................... | 7 | 393 | 400 |

Expression (6) gives the probability for Table 7.5 as

$$\frac{7!393!200!200!}{400!0!7!193!200!} = 0.0074 \tag{8}$$

The sum of the three probabilities (5), (7), and (8), or 0.2244, is the probability that the experiment will show as much or more apparent superiority of the experimental method as Table 7.2 shows, if in fact there is no real difference. It should be noted that we add together the probabilities of only those cases representing greater superiority for the experimental method than was observed in the sample. We do not include the probabilities of cases representing greater superiority for the standard than was observed for the experimental method. This is because we are analyzing the data in order to decide whether the experimental is superior to the standard method, not simply to decide whether there is a difference between the methods without regard to its direction.

Finally, a conclusion is reached regarding the superiority of the experimental method as follows: We agree to accept the experimental method as superior to the standard on the basis of an analysis of the above type only when the resulting probability is less than $\alpha$, where $\alpha$ is chosen in advance and is generally taken to be 0.10, 0.05, 0.01, or 0.001, depending upon the seriousness of accepting the experimental method as superior when it is not. Then $\alpha$ is the probability of accepting the experimental method as superior when actually there is no real difference between the methods. In terms of any of the usual values of $\alpha$, the probability 0.2244 would imply that the data of Table 7.2 provide no basis for concluding that the experimental method is actually superior to the standard.

## 2.2. Short-cut Methods

The evaluation of the factorials involved in expressions (5), (7), and (8) was done by cancellation of common factors. This procedure is the simplest in this case, but in general it may be very laborious. The calculations can be considerably reduced by using a table of the logarithms of factorials.*

A much simpler procedure, which usually yields very satisfactory approximations to the exact method described in Sec. 2.1, is available when at least two (and preferably five) successes would be expected in each cell if there were no difference in the long-run probabilities for the two methods. This expected number of successes is computed for any cell by dividing the product of the marginal totals for that cell by the total number of trials made; thus, the expected number of successes for the standard method is, on the basis of Table 7.3,

$$\frac{(a + b)(a + c)}{a + b + c + d} \tag{9}$$

* A convenient table is given in Fry, *Probability and Its Engineering Uses.* Shorter tables can be found in various handbooks.

In such cases we calculate a quantity $\chi_c$ as follows:

First, adjust the table of observations so that each frequency is brought ½ unit nearer the frequency expected if there were no long-run difference between the two methods. This may be done by adjusting one frequency and then changing the others so that the marginal totals of the new table equal those of the old. For example, in Table 7.2 the expected frequency for the first cell is $\dfrac{7 \times 200}{400} = 3\frac{1}{2}$, so we increase the observed figure of 2 to 2½. Changing the other figures to keep the marginal totals fixed, we have

TABLE 7.6

|  | Successes | Failures | Total |
|---|---|---|---|
| Standard........................................ | 2½ | 197½ | 200 |
| Experimental................................... | 4½ | 195½ | 200 |
| Total......................................... | 7 | 393 | 400 |

Table 7.6 represents Table 7.2 "corrected for continuity."*

Second, each adjusted frequency is subtracted from the corresponding expected frequency, the differences are squared, the squares are divided by the corresponding expected frequencies, and the ratios are summed to obtain $\chi_c^2$.

The final value of $\chi_c^2$ may be obtained directly without going through the intermediate steps, as follows: Calculate the products of frequencies diagonally opposite each other in the adjusted table; subtract one diagonal product from the other; square the difference; multiply the square by the sum of the four frequencies in the table; and divide this product by the product of all four marginal totals. From the data of Table 7.6, we have

$$\chi_c^2 = \frac{(2\frac{1}{2} \times 195\frac{1}{2} - 4\frac{1}{2} \times 197\frac{1}{2})^2 \times 400}{7 \times 393 \times 200 \times 200} = 0.5816 \qquad (10)$$

More generally, if we replace $a$, $b$, $c$, and $d$ in Table 7.3 by $a'$, $b'$, $c'$, and $d'$, representing the frequencies corrected for continuity, (10) becomes

$$\chi_c^2 = \frac{(a'd' - b'c')^2(a + b + c + d)}{(a + b)(a + c)(c + d)(b + d)} \qquad (11)$$

---

* Although this account of the continuity correction describes the procedure accurately, it may convey an erroneous idea of the theoretical basis for the correction. The general procedure is to carry through the calculation of $\chi$ from the original data, then find the next lowest value of $\chi$ that could have arisen, and finally take $\chi_c$ as halfway between these two values. In the particular case of double dichotomies, this value of $\chi$ can be found most easily by correcting the original data as in Table 7.6; but in most other cases where the $\chi$ test is used, correction of the original data will lead to serious error in the value of $\chi$. For an excellent analysis and presentation of this point, see Cochran, "The $\chi^2$ Correction for Continuity," especially p. 423.

Next, we take the square root and find $\chi_c = 0.7626$. This may be treated as a normal deviate; that is, the required probability is given by

$$\frac{1}{\sqrt{2\pi}}\int_{\chi_c}^{\infty} e^{-\frac{1}{2}t^2}\, dt \tag{12}$$

Referring to a table of the normal probability integral,* for example Table 1.5 of Chap. 1, we find the probability to be 0.2228. This compares favorably with the exact probability of 0.2244 found in Sec. 2.1.

More accurate values of $\chi_c$ corresponding to the probabilities 0.025 and 0.005 than are implied by (12) are given in Table VIII, "Tests of Significance for $2 \times 2$ Contingency Tables," in Fisher and Yates, *Statistical Tables*. Detailed instructions and illustrations for the use of this table are given in its introduction.

When the percentage of successes is small and the number of trials in each series large, as is the case in the examples discussed here, the successes may be analyzed by the binomial distribution, along the lines discussed in Sec. 1 and the second paragraph of Sec. 2.1.† The value of $P$ for the binomial test is the proportion of all trials that fall into, say, the standard series. In the illustration of Table 7.2, $P = \frac{1}{2}$. According to the binomial distribution, the probability that seven trials of an event having two equally likely outcomes will show as many as five or more of a designated type (in other words, the probability that with seven successes observed five or more will occur in the experimental series) is

$$\frac{1}{2^7}\left(\frac{7!}{2!5!} + \frac{7!}{1!6!} + \frac{7!}{0!7!}\right) = 0.2266 \tag{13}$$

This conforms satisfactorily to the exact value 0.2244 found in Sec. 2.1.

### 2.3.  Alternative Methods

A procedure that has been widely used for testing the significance of the difference between two proportions is based on the following considerations: The difference between the two percentages is divided by the standard error of the difference, and the ratio is regarded as normally distributed. The standard error of the difference is computed from the familiar relation that the standard error of the difference between two independent statistical variables is the square root of the sum of the squared standard deviations of the two variables. The standard error of a proportion $p$ is well known to be $\sqrt{P(1-P)/N}$,

---

* For a detailed table, see the tables prepared by the New York Mathematical Tables Project.

† See PRZYBOROWSKI and WILEŃSKI, "Homogeneity of Results in Testing Samples from Poisson Series," especially Sec. 2, Derivation of Test, pp. 314–317.

where $P$ is the true population value of the proportion and $N$ is the number of trials.

Applying these considerations, it is often concluded that

$$V = \frac{\sqrt{N}(p_E - p_S)}{\sqrt{p_S(1 - p_S) + p_E(1 - p_E)}} \tag{14}$$

where $p_E$ is the observed proportion for the experimental method

$p_S$ is the observed proportion for the standard method

$N$ is the number of trials with *each* method

may be treated as normally distributed with mean of 0 and standard deviation of 1. If the probability of obtaining a normal deviate as large as $V$ is small, it is concluded that the difference in the proportions is greater than can be ascribed to chance.

There are two serious objections to the procedure by which expression (14) is derived: (1) $V$ will not actually be distributed as assumed; and (2) the procedure is not self-consistent.

With regard to the distribution of $V$, there are two considerations: In the first place, the true population standard deviation of the difference is required in the denominator if the ratio is to be normally distributed with unit variance when $N$ is large; but actually only experimental values are available. Hence, chance variations affect not only the numerator but also the denominator of $V$. In the second place, even if we know the true standard deviation, the distribution of the difference $(p_E - p_S)$ cannot be satisfactorily approximated by a normal distribution when the proportions are near 0 or 1, unless $N$ is very large.

The procedure leading to expression (14) is inconsistent in two respects. It sets out to calculate the probability that, if there were actually no long-run difference, a difference as great as that observed would occur by chance. But if there were no long-run difference in the proportions, there would be none in the standard deviations. Therefore, when the standard deviation is estimated from the experiment, the estimate should be based on the two series together. This leads to considering

$$\chi = \frac{\sqrt{N}\,(p_E - p_S)}{\sqrt{(p_S + p_E)\left(1 - \dfrac{p_S + p_E}{2}\right)}} \tag{15}$$

as a normally distributed variable.

The second inconsistency in expression (14) may be seen as follows: It would be perfectly logical to find what proportion of the total successes were made with the standard method and what proportion of the total failures were made with the experimental method. These proportions should be equal if the standard and experimental methods are

equally effective. But if the significance of the difference of this pair of proportions is tested by expression (14), the result is not the same as when the logically equivalent pair based on methods is tested. Expression (15), however, is consistent in this respect as well as in the standard error it uses.

Expression (15) is mathematically identical with formula (11) in the previous section, except that formula (11) is calculated from the data corrected for continuity. Thus, in terms of the frequencies of Table 7.3 (without continuity corrections), expression (15) can be written as

$$\chi = \frac{\sqrt{a+b+c+d}\,(ad-bc)}{\sqrt{(a+b)(a+c)(b+d)(c+d)}} \tag{16}$$

with $a + b = c + d = N$. This is simply the square root of expression (11), neglecting the continuity corrections. When the continuity corrections are taken into account and $N$ is introduced for simplification, expression (16) becomes

$$\chi_c = \frac{ad-bc+N}{\sqrt{(N/8)(a+c)(b+d)}} \tag{17}$$

Note that $N$ in expressions (14), (15), (16), and (17) represents the number of trials in *each* set. Also note that (17) is valid only when the continuity corrections increase $a$ and $d$ and decrease $b$ and $c$; for the reverse situation, $+N$ in the numerator of (17) must be replaced by $-N$.

Expression (14), though it is theoretically inferior to expressions (15) and (17), does not differ materially from them when $N$ is large. Since (15) and (17) are theoretically superior and are little or no more trouble to calculate, they are always preferable to (14). Even (15) and (17), however, are approximations; they fail for very small values of $N$ or very small probabilities of a success, and in such cases the procedures outlined above in Secs. 2.1 and 2.2 must be used. The procedure of Secs. 2.1 and 2.2 should be used also when $a + b$ is different from $c + d$.

For practical work with double dichotomies, the Zubin Item Analyzer gives results essentially equivalent to those of expression (14).* In this device the difficulties that (1) proportions are not normally distributed when the true proportion is small, and (2) the standard deviation depends upon the true proportion, are avoided by employing a new variable $\theta$, which is a function of $p$. The function is chosen so that, first, $\theta$ is more nearly normally distributed than $p$, and second, the standard deviation of $\theta$ is independent of the true long-run value of $\theta$ or $p$. A

---

* For a description of this device, see Fulcher and Zubin, "The Item Analyzer." A model constructed by Dr. Fulcher was lent to the Statistical Research Group for trial in 1943. It proved quick and easy to use and would be cheap to build, as it involves only plywood and celluloid.

function that meets these requirements when $N$ is not too small is

$$\theta = 2 \arcsin \sqrt{p} \qquad (18)$$

Unless $N$ is very small, $\theta$ may be regarded as normally distributed; the standard deviation is $1/\sqrt{N}$ when $\theta$ is measured in radians or $57.3/\sqrt{N}$ when $\theta$ is measured in degrees. The Zubin Analyzer is a mechanical device for transforming the proportions to the proper angles and reading directly the difference $(\theta_E - \theta_S)$. The required probability can then be read from a simple chart showing the difference and its standard error, $\sqrt{2/N}$.*

### 3. NECESSARY NUMBER OF TRIALS

A method that is sometimes recommended for finding the required sample size in experiments involving double dichotomies is to set $V$ in expression (14) equal to the value $V_\alpha$, which would be exceeded only $\alpha$ percent of the time if there were no true differences in the proportions, and then solve for $N$, using values of $p_S$ and $p_E$ obtained experimentally. The result is

$$N = \frac{V_\alpha^2[p_S(1 - p_S) + p_E(1 - p_E)]}{(p_E - p_S)^2}, \qquad (19)$$

This procedure can be made logically consistent by basing the calculations on expression (15) instead of on (14), in which case the formula becomes

$$N = \frac{K_\alpha^2(p_S + p_E)\left(1 - \dfrac{p_S + p_E}{2}\right)}{(p_E - p_S)^2} \qquad (20)$$

where $K_\alpha$ is the normal deviate exceeded with probability $\alpha$, as given in Table 1.1 of Chap. 1.†

In general, (20) solves correctly the following problem: If the proportion of successes is $p_S$ for the standard and $p_E$ for the experimental method, how large must be the number of trials from which these proportions were computed before the experiment can be regarded as indicative of a real difference? The question stated is, however, of little or no practical significance. The practical problem exists only because it cannot be assumed that the proportions shown in the experiment are

---

* See Chap. 16 for an account of the inverse sine transformation for percentages.

† The value of $K_\alpha^2$ can be found from tables of the chi-square distribution, for example, Table IV in Fisher and Yates, *Statistical Tables*, thus eliminating the need of squaring $K_\alpha$. Since the probability that $\chi^2$ will exceed a certain level is the probability that $\chi$ will exceed the square root of that level *plus* the probability that $-\chi$ will exceed the square root of that level, the probabilities given in the $\chi^2$ tables must be halved for the present purposes.

the true proportions. If they were, we should be able to decide without difficulty which method is superior. Furthermore, even if we were willing to accept the proportions $p_s$ and $p_E$ shown in the experiment as suitable approximations of the true long-run proportions $P_S$ and $P_E$, we could compute $N$ from (20) only after carrying out the experiment to find $p_s$ and $p_E$.

To determine a useful formula for $N$, we must think of formula (11) or its equivalent, formula (15), as constituting a rule according to which we decide from the data which of two actions to take. If the formula yields a value above some predetermined critical level, we accept the experimental method as superior, and if it yields a lower value we do not accept it (that is, for practical purposes we reject it). If the true proportion of successes for the experimental method is the same as the true proportion of successes for the standard method, it is important that the critical level of $\chi$ be such that there is only a small probability, say $\alpha$, that the data will lead to a value of $\chi$ above the critical level. Similarly, if the true proportion of successes is $P_s$ for the standard method but $P_E$ $(P_E > P_s)$ for the experimental method, it is important that the critical level of $\chi$ and the sample size $N$ be such that there is only a small probability, say $\beta$, that the data will lead to a value of $\chi$ below the critical level.

The first condition, that the probability be only $\alpha$ of accepting the experimental method as superior when the true probabilities of success are equal, is met simply by selecting the critical level of $\chi$ as $K_\alpha$, the normal deviate exceeded with probability $\alpha$. Then, regardless of the sample size, the probability is $\alpha$ that the experimental method will be accepted when the true probabilities of success are the same for both methods.

On the other hand, for small samples the probability that $\chi$ will exceed $K_\alpha$ will not be much more than $\alpha$ even when the true probabilities are $P_s$ and $P_E$. This probability of accepting the experimental method when the true probabilities are $P_s$ and $P_E$ will, however, increase as the sample size increases; and the problem is to find that value of $N$ for which it is $1 - \beta$.

The required value of $N$, the size of the sample to be used with *each* method, is given by

$$N = \frac{1}{2}\left(\frac{K_\alpha + K_\beta}{\arcsin \sqrt{P_s} - \arcsin \sqrt{P_E}}\right)^2 \qquad (21)$$

if angles are measured in radians, or by

$$N = 1641.6\left(\frac{K_\alpha + K_\beta}{\arcsin \sqrt{P_s} - \arcsin \sqrt{P_E}}\right)^2 \qquad (22)$$

if angles are measured in degrees, where $K_\epsilon$ is defined by

$$\frac{1}{\sqrt{2\pi}} \int_{K_\epsilon}^{\infty} e^{-\frac{x^2}{2}} \, dx = \epsilon \qquad (23)$$

$\alpha$ is the probability of an error of the "first kind," that is, of accepting the experimental method as superior when actually the probability of success is the same for both methods

$\beta$ is the probability of an error of the "second kind," that is, of rejecting the experimental method when the true probabilities for the standard and experimental methods are $P_S$ and $P_E$, respectively.*

Values of $N$ can be read directly from the nomogram of Fig. 7.1.

The choice of $P_S$ can usually be based upon previous experience with the standard method, which will indicate approximately its proportion of successes. The value of $P_E$ may then be chosen enough above $P_S$ so that the difference is one that it is important to detect when present. Moderate variations in the values of $P_S$ and $P_E$ are not serious if $P_S$ and $P_E$ are affected alike. If the angular difference, the square of which constitutes the denominator of formula (22), is accepted as the appropriate measure of the difference between two percentages, it is not necessary to choose $P_S$ and $P_E$ individually but only to determine the angular difference that should be detected.†

To illustrate the application of formula (21), suppose $P_S = 0.010$, $P_E = 0.025$, $\alpha = 0.05$, and $\beta = 0.10$. From Table 1.1 of Chap. 1, we find $K_\alpha = 1.645$ and $K_\beta = 1.282$. From trigonometric tables,‡ we find arcsin 0.1000 and arcsin 0.1581 to be, respectively, 0.1002 and 0.1588 radians. Substituting these figures into (18), we have

$$N = \frac{(1.645 + 1.282)^2}{2(0.1588 - 0.1002)^2} \qquad (24)$$
$$= 1,246$$

* If only one method is involved in an experiment, and if it is desired that the probability of approving it be $\alpha$ or $1 - \beta$ according as its true probability of success is $P_S$ or $P_E$, formula (21) will indicate twice the appropriate sample size; that is, the appropriate sample size is given by substituting 4 for 2 in the denominator of formula (21). This was the method of computing the sample sizes $N_a$ given in Table 1.3 of Chap. 1 for acceptance-inspection plans based on attributes. For that application, the first sentence of this footnote should be changed to read as follows: If only one product is involved in an inspection plan, and if it is desired that the probability of rejecting be $\alpha$ or $1 - \beta$ according as the true probability of a defective is $p_1$ or $p_2$, formula (21) will indicate twice the appropriate sample size.

† In this connection, see Statistical Research Group, *Sequential Analysis*, Sec. 3.62, pp. 3.47–3.50, especially Table 3.61 and Fig. 3.61; also Secs. 3.21 and 3.22, pp. 3.09–3.10.

‡ For example, Mathematical Tables Project, *Arc sin x.*

# NUMBER OF CASES REQUIRED FOR COMPARING TWO PERCENTAGES

## Directions

1. Select values of $\alpha$, $\beta$, $P_s$ and $P_e$ (See "Explanation").
2. Locate the value of $\alpha$ on the scale marked "$\alpha$."
3. Locate the value of $\beta$ on the scale marked "$\beta$."
4. Locate the point at which a straight line from $\alpha$ to $\beta$ cuts the scale marked "$A$."
5. Locate the value of $P_s$ on the scale marked "$P_s$." If $P_s$ exceeds 0.50 use $1 - P_s$.
6. Locate the value of $P_e$ to 1 and make a corresponding change of $P_s$ to $1 - P_e$.
7. Locate the point of $P_s$ on the scale marked "$P_s$."
8. Locate the point at which a straight line from $P_s$ to $P_e$ cuts the scale marked "$N$." The "$N$" is cut by a straight line from the point on "$B$" to the point on "$A$." The "$N$" scale is calibrated so that the required sample size $N$ is given opposite this point.
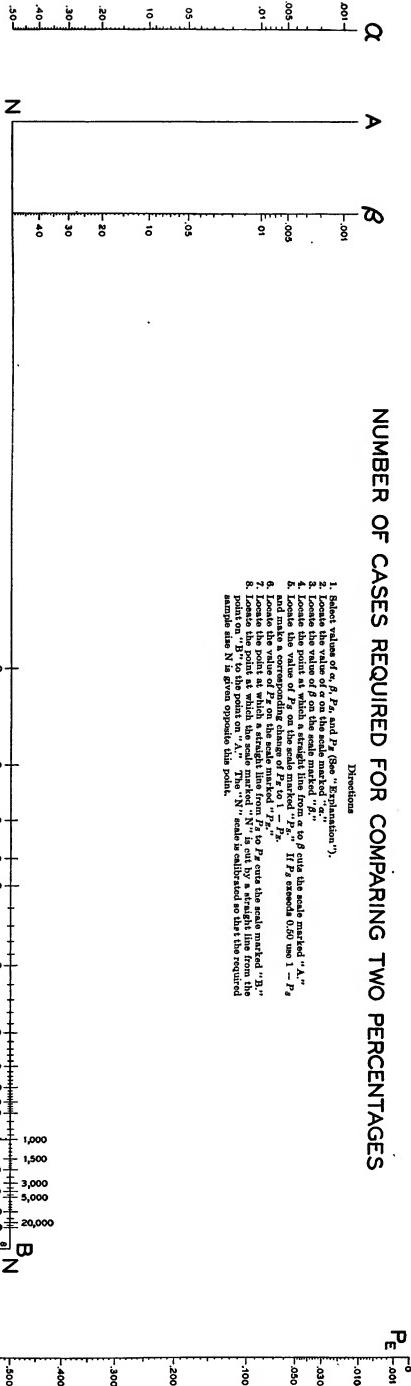
## Explanation

Call the two percentages being compared the "standard" and the "experimental."

$\alpha$ is defined as follows: We will conclude that the experimental method is better only if it is above in the experiment a superiority so great that the probability of this much superiority would be only $\alpha$ if the true long-run percentages were equal. In other words, $\alpha$ is the probability that we will erroneously judge the experimental method better when in fact the two methods are equally good.

$\beta$ is the probability that we will fail to judge the experimental method better when in fact it is better. In other words, it is the probability that the experimental method, although actually better in the long run, will fail to show a smaller superiority in the experiment than is required by the value of $\alpha$ for concluding that it is better.

$P_s$ is an estimate of the true long-run probability for the standard method.

$P_e$ is an estimate of the true long-run probability for the experimental method. This estimate may be based on prior experience, scientific judgment, etc. This estimate may be determined like that for $P_s$; or it may be the smallest figure above $P_s$ for which it would be important to detect the superiority. For example, if $P_s = 0.10$, and if it would be important to detect any superiority of more than 30 percent, we set $P_e = 0.13$.

$N$ is the number of trials with each method required in the experiment, as given by the formula $N = \dfrac{1}{2}\left(\dfrac{K_\alpha + K_\beta}{\arcsin \sqrt{P_e} - \arcsin \sqrt{P_s}}\right)^2$

where $K_\alpha$ is given by:

$$\frac{1}{\sqrt{2\pi}} \int_{K_\alpha}^{\infty} e^{-\frac{x^2}{2}}\, dx = \epsilon$$

and angles are measured in radians.

Fig. 7.1.—Number of cases required for comparing two percentages.

if angles are measured in degrees, where $K_\epsilon$ is defined by

$$\frac{1}{\sqrt{2\pi}} \int_{K_\epsilon}^{\infty} e^{-\frac{x^2}{2}} dx = \epsilon \tag{23}$$

$\alpha$ is the probability of an error of the "first kind," that is, of accepting the experimental method as superior when actually the probability of success is the same for both methods

$\beta$ is the probability of an error of the "second kind," that is, of rejecting the experimental method when the true probabilities for the standard and experimental methods are $P_S$ and $P_E$, respectively.*

Values of $N$ can be read directly from the nomogram of Fig. 7.1.

The choice of $P_S$ can usually be based upon previous experience with the standard method, which will indicate approximately its proportion of successes. The value of $P_E$ may then be chosen enough above $P_S$ so that the difference is one that it is important to detect when present. Moderate variations in the values of $P_S$ and $P_E$ are not serious if $P_S$ and $P_E$ are affected alike. If the angular difference, the square of which constitutes the denominator of formula (22), is accepted as the appropriate measure of the difference between two percentages, it is not necessary to choose $P_S$ and $P_E$ individually but only to determine the angular difference that should be detected.†

To illustrate the application of formula (21), suppose $P_S = 0.010$, $P_E = 0.025$, $\alpha = 0.05$, and $\beta = 0.10$. From Table 1.1 of Chap. 1, we find $K_\alpha = 1.645$ and $K_\beta = 1.282$. From trigonometric tables,‡ we find arcsin 0.1000 and arcsin 0.1581 to be, respectively, 0.1002 and 0.1588 radians. Substituting these figures into (18), we have

$$N = \frac{(1.645 + 1.282)^2}{2(0.1588 - 0.1002)^2} \tag{24}$$
$$= 1{,}246$$

---

* If only one method is involved in an experiment, and if it is desired that the probability of approving it be $\alpha$ or $1 - \beta$ according as its true probability of success is $P_S$ or $P_E$, formula (21) will indicate twice the appropriate sample size; that is, the appropriate sample size is given by substituting 4 for 2 in the denominator of formula (21). This was the method of computing the sample sizes $N_a$ given in Table 1.3 of Chap. 1 for acceptance-inspection plans based on attributes. For that application, the first sentence of this footnote should be changed to read as follows: If only one product is involved in an inspection plan, and if it is desired that the probability of rejecting be $\alpha$ or $1 - \beta$ according as the true probability of a defective is $p_1$ or $p_2$, formula (21) will indicate twice the appropriate sample size.

† In this connection, see Statistical Research Group, *Sequential Analysis*, Sec. 3.62, pp. 3.47–3.50, especially Table 3.61 and Fig. 3.61; also Secs. 3.21 and 3.22, pp. 3.09–3.16.

‡ For example, Mathematical Tables Project, *Arc sin x*.

If the nomogram is used, we pass a line between 0.05 on the $\alpha$ scale and 0.10 on the $\beta$ scale and mark its point of intersection with the $A$ scale. Another line is passed between 0.01 on the $P_S$ scale and 0.025 on the $P_E$ scale, and the point of intersection of this line with the $B$ scale is marked. The $N$ scale is calibrated so that the number of trials needed for each method can be read off where the $N$ scale is intersected by a straight line between the points marked on the $A$ and $B$ scales. In this example, the point of intersection falls about halfway between $N = 1{,}200$ and $N = 1{,}300$.

## 4. OPERATING CHARACTERISTICS

Conclusions drawn from double dichotomies depend upon two quantities, one characterizing the amount of data collected and the other characterizing the method of analyzing the data. These quantities are $N$, the number of observations with each method, and $K_\alpha$, the critical level of $\chi$ which, if equaled or exceeded, leads to acceptance of the experimental method as superior. Formula (21) can be used to determine the operating characteristics of any plan for analyzing a double dichotomy defined by given values of $N$ and $K_\alpha$.

By the operating characteristics of a plan is meant the probability, as a function of the true nature of the populations, that the plan will lead to accepting the experimental method as superior. In this case, the operating characteristics may be represented by a surface rather than a curve, for the true nature of the populations is measured by two parameters $P_S$ and $P_E$. If in the plane of the page we measure $P_E$ lengthwise and $P_S$ crosswise, any point for which $P_E$ and $P_S$ are both between 0 and 1 represents a possible combination of $P_E$ and $P_S$. Vertically above each point we may erect a line whose height represents the probability that a test using the given $N$ and $K_\alpha$ will lead to accepting the experimental method as superior. The surface formed by the tops of these lines is the OC surface. A test is satisfactory if its OC surface shows high probabilities of accepting the experimental method when its superiority is sufficient to make such acceptance important, and low probabilities of accepting the experimental when its superiority is so small (especially when it is zero or negative) as to make its acceptance undesirable.

We may rewrite (21) as

$$-K_\beta = K_\alpha - \sqrt{2N}\,(\arcsin\sqrt{P_E} - \arcsin\sqrt{P_S}) \qquad (25)$$

Formula (25) yields $-K_\beta$; the probability of accepting the experimental method is $1 - \beta$, which is the probability that a normal deviate will exceed $-K_\beta$ and may be obtained from Table 1.5 of Chap. 1. Thus, for any combination of $P_E$ and $P_S$, formula (25) yields the probability that a test based on $N$ and $K_\alpha$ will lead to acceptance of the experimental method.

Formula (25) does not depend upon the individual probabilities $P_E$ and $P_S$ but only upon their angular difference

$$T = \arcsin \sqrt{P_E} - \arcsin \sqrt{P_S} \qquad (26)$$

so we may write it

$$K_\beta = \sqrt{2N}\, T - K_\alpha \qquad (27)$$

Letting $L_T$ denote the probability that a test defined by $N$ and $K_\alpha$ will lead to acceptance of the experimental method when the true angular difference is $T$, and noting that $K_{1-\beta} = -K_\beta$, we have

$$L_T = \frac{1}{\sqrt{2\pi}} \int_{K_\alpha - \sqrt{2N}\, T}^{\infty} e^{-\frac{x^2}{2}}\, dx \qquad (28)$$

as the equation of the OC curve in terms of $T$.   Values of $L_T$ can be obtained from Table 1.5, Chap. 1, by taking

$$K_\epsilon = K_\alpha - \sqrt{2N}\, T \qquad (29)$$
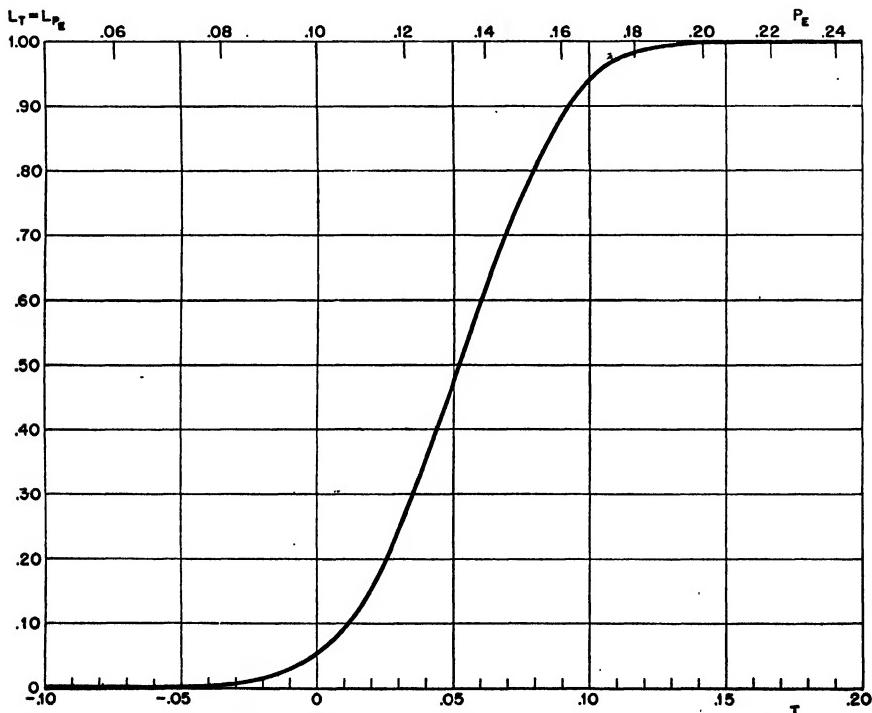


FIG. 7.2.—Operating-characteristic curve for a comparison of two percentages, $N = 500$, $K_\alpha = 1.645$.

Figure 7.1 can be used to determine $L_T$ from (28). The given values of $P_E$ and $P_S$ determine a point on the $B$ scale. A line is drawn through the point on the $B$ scale and the given value of $N$ on the $N$ scale and extended until it intersects the $A$ scale. This point on the $A$ scale and the given value of $\alpha$ on the $\alpha$ scale determine a line whose intersection with the $\beta$ scale gives $1 - L_T$.

In evaluating a proposed test, it is sometimes helpful to plot the OC curve against the true proportion of successes for the experimental method, assuming some value, say $P'_S$, for the proportion of successes with the standard method. This gives a picture, the significance of which depends on the validity of the value used for $P'_S$, of the absolute level of performance by the experimental method that will result in a high probability of its acceptance or rejection. Such a picture can be obtained from formula (25), replacing $P_S$ by $P'_S$ and calculating $L_T = 1 - \beta$ for various values of $P_E$.

Figure 7.2 shows the OC curve for an analysis based on a sample size $N$ of 500 for each method and a critical level of $\chi$, $K_\alpha$, of 1.645. Two horizontal scales are shown, the lower in terms of $T$, the true angular difference between the two probabilities, and the upper in terms of $P_E$ on the assumption that $P_S = 0.10$.

## 5. TECHNICAL NOTES

Let $s_E$ and $s_S$ be the observed numbers of successes in $N$ trials with the experimental and standard methods, respectively. Then the observed proportions of successes with the two methods are

$$p_E = \frac{s_E}{N} \qquad \text{and} \qquad p_S = \frac{s_S}{N} \qquad (30)$$

When we are interested only in alternative hypotheses for which $P_E$, the true proportion of successes with the experimental method, is greater than $P_S$, the long-run proportion of successes with the standard method, the classical test with level of significance $\alpha$ is to reject the null hypothesis that $P_E = P_S$ if

$$\frac{\sqrt{N}(p_E - p_S)}{\sqrt{(p_E + p_S)\left(1 - \frac{p_E + p_S}{2}\right)}} \geq K_\alpha \qquad (31)$$

where

$$\frac{1}{\sqrt{2\pi}} \int_{K_\alpha}^{\infty} e^{-\frac{x^2}{2}} dx = \alpha \qquad (32)$$

Except when $N$ is very small, or $P_E$ or $P_S$ very close to 0 or 1, this test is essentially equivalent to the criterion: Reject the null hypothesis if *

$$\sqrt{2N}\,(\arcsin \sqrt{p_E} - \arcsin \sqrt{p_S}) \geq K_\alpha \qquad (33)$$

* See Chap. 16 for a discussion of the inverse sine transformation of percentages, upon which what follows is based.

because the quantity

$$t = \arcsin \sqrt{p_E} - \arcsin \sqrt{p_S} \tag{34}$$

may be regarded as normally distributed with mean of

$$T = \arcsin \sqrt{P_E} - \arcsin \sqrt{P_S} \tag{35}$$

and variance $1/2N$, whence it follows that when the null hypothesis is true, $\sqrt{2N}\, t$ is normally distributed about zero with unit standard deviation.

Now, to find $N$ so that the probability is $\beta$ of accepting the null hypothesis when $T = T'$, it is necessary that

$$\text{Prob } [\sqrt{2N}\, (\arcsin \sqrt{p_E} - \arcsin \sqrt{p_S}) < K_\alpha | T'] = \beta \tag{36}$$

This is equivalent to requiring that

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{K_\alpha - \sqrt{2N} T'} e^{-\frac{x^2}{2}} \, dx = \beta \tag{37}$$

which becomes

$$K_\alpha - \sqrt{2N}\, T' = -K_\beta \tag{38}$$

where $K_\beta$ is defined by replacing $\alpha$ by $\beta$ in equation (32). Solving for $N$ gives

$$N = \frac{1}{2}\left(\frac{K_\alpha + K_\beta}{T'}\right)^2 \tag{39}$$

or

$$N = \frac{1}{2}\left(\frac{K_\alpha + K_\beta}{\arcsin \sqrt{P_E} - \arcsin \sqrt{P_S}}\right)^2 \tag{40}$$

It may be observed that formula (40) follows immediately from formula (148) in Chap. 1, Sec. 11.1.3. We have

$$Z_1 = 0 \tag{41}$$

$$Z_2 = \arcsin \sqrt{P_E} - \arcsin \sqrt{P_S} \tag{42}$$

$$\sigma = 1 \tag{43}$$

$$F(N) = \frac{1}{\sqrt{2N}} \tag{44}$$

Substituting into equation (148) of Chap. 1, we have

$$\frac{1}{\sqrt{2N}} = \frac{\arcsin \sqrt{P_E} - \arcsin \sqrt{P_S}}{K_\alpha + K_\beta} \tag{45}$$

which, when solved for $N$, leads to (40). Similarly, equation (147) of Chap. 1 leads to a critical level $A$ of $K_\alpha/\sqrt{2N}$ for $\arcsin \sqrt{p_E}$ — $\arcsin \sqrt{p_S}$, which is equivalent to (33).

## 6. REFERENCES

COCHRAN, W. G.: "The $\chi^2$ Correction for Continuity," *Iowa State College Journal of Science*, Vol. 16 (1942), pp. 421–436.

FISHER, R. A., *Statistical Methods for Research Workers*, 5th and later ed., Oliver and Boyd, Ltd., London and Edinburgh, 1934–44.

—— and YATES, FRANK: *Statistical Tables for Biological, Agricultural and Medical Research*, 2d ed., Oliver and Boyd, Ltd., London and Edinburgh, 1943, 98 + viii pp.

FRY, THORNTON C.: *Probability and Its Engineering Uses*, D. Van Nostrand Company, New York, 1926.

FULCHER, JOHN S., and ZUBIN, JOSEPH: "The Item Analyzer: A Mechanical Device for Treating the Four Fold Table in Large Samples," *Journal of Applied Psychology*, Vol. 26 (1942), pp. 511–522.

Mathematical Tables Project, Work Projects Administration for the City of New York, Federal Works Agency, *Tables of Probability Functions*, Vol. 2, National Bureau of Standards, Washington, D.C., 1942, 344 + xxi pp.

——*Table of Arc sin x*, Columbia University Press, New York, 1945, 121 + xix pp.

PRZYBOROWSKI, J., and WILEŃSKI, H.: "Homogeneity of Results in Testing Samples from Poisson Series, with an Application to Testing Clover Seed for Dodder," *Biometrika*, Vol. 31 (1940), pp. 313–323.

# CHAPTER 8

# PLANNING AND INTERPRETING EXPERIMENTS FOR COMPARING TWO STANDARD DEVIATIONS

## FIGURES

## TABLES

This chapter was prepared by
Churchill Eisenhart, National Bureau of Standards

# PLANNING AND INTERPRETING EXPERIMENTS FOR COMPARING TWO STANDARD DEVIATIONS

## 1. INTRODUCTION

The present chapter is concerned with the planning and interpretation of experiments to determine whether (1) a given product, or process, is more (or less) variable than a standard product, or process; (2) one of two new products, or processes, is more (or less) variable than the other; (3) the random variation of a product, or process, has a component assignable to a specific source of variation.

With regard to proposed experiments, the problem considered is that of determining how many observations will be required in order to render the discriminating power of the experiment adequate for the purposes at hand. The solutions to this problem, in addition to being useful in the planning of experiments, provide a means of evaluating the discriminating power of an experiment already carried out. Thus, in planning experiments for the three purposes listed above, the results given below will enable a research worker to formulate an experiment the size of which, on the one hand, is compatible with the cost (monetary expense, time, inconvenience, and so forth) of obtaining the observations and, on the other, limits to a known extent the risks of making erroneous decisions of practical significance. Alternatively, in the case of experiments already conducted, the results given below will enable the experimenter to judge the confidence that he can place in conclusions based on the data obtained.

## 2. EXPERIMENTS TO DETERMINE WHETHER A NEW PRODUCT IS MORE VARIABLE THAN A STANDARD PRODUCT

### 2.1. Problem

The following situation is common in practical work: A certain product has been in use for some time and its variability with respect to some important measurable characteristic $x$ is known. Let $\sigma_0^2$ denote the known variance (squared standard deviation) of $x$ for this standard product. A new product is under consideration, whose variability with respect to $x$ is unknown; let $\sigma_1^2$ denote the unknown variance of $x$ for the new product. Since the value of $\sigma_1^2$ is unknown, it is not possible to

decide with certainty whether the variability of the new product is more than that of the standard, that is, whether $\sigma_1^2/\sigma_0^2 > 1$.

### 2.2.  General Procedure for Basing an Answer on Observational Data

An answer to this question can, however, be obtained by computing an estimate $s_1^2$ of $\sigma_1^2$ from values of $x$ observed in a sample of the new product and agreeing that $s_1^2/\sigma_0^2 > C$ will be taken to imply that $\sigma_1^2/\sigma_0^2 > 1$, and $s_1^2/\sigma_0^2 \leq C$ will be taken to imply that $\sigma_1^2/\sigma_0^2 \leq 1$, where $C$ is some positive constant chosen in advance, that is, before the value of $s_1^2$ has been obtained, preferably before the experiment determining the value of $s_1^2$ is carried out.  It is evident that if $C$ is very large, application of this procedure will almost always lead to the decision that the new product is not more variable than the standard ($\sigma_1^2 \leq \sigma_0^2$); if $C$ is very small, the decision will almost always be that the new is more variable than the standard ($\sigma_1^2 > \sigma_0^2$).  Clearly, some intermediate value of $C$ is called for—a value that will not lead too frequently to the decision $\sigma_1^2/\sigma_0^2 > 1$ when the reverse is true and yet will lead to this decision in most of the cases in which the excess of $\sigma_1^2$ over $\sigma_0^2$ is of practical importance.

### 2.3.  Standard Procedure*

The standard procedure is to use as an estimate of $\sigma_1^2$ a function $s_1^2$ of the observed values of $x$ that has the property that in repeated sampling $n_1 s_1^2/\sigma_1^2$ is distributed, at least to a good approximation, as $\chi^2$ for $n$ degrees of freedom, and to

a.  Conclude $\sigma_1^2/\sigma_0^2 \leq 1$ whenever

$$\frac{s_1^2}{\sigma_0^2} \leq \frac{\chi_\alpha^2(n_1)}{n_1} \tag{1}$$

b.  Conclude $\sigma_1^2/\sigma_0^2 > 1$ whenever

$$\frac{s_1^2}{\sigma_0^2} > \frac{\chi_\alpha^2(n_1)}{n_1} \tag{2}$$

where $\chi_\alpha^2(n_1)$ is the $\alpha$ level of significance of $\chi^2$ for $n_1$ degrees of freedom, that is

$$\alpha = P\{\chi^2(n_1) > \chi_\alpha^2(n_1)\} = \frac{1}{\Gamma\left(\dfrac{n}{2}\right)} \int_{\chi_{\alpha(n_1)}^2}^{\infty} \left(\frac{1}{2}x^2\right)^{\frac{1}{2}(n-2)} e^{-\frac{1}{2}x^2} d\left(\frac{1}{2}x^2\right) \tag{3}$$

for $n = n_1$.†

---

* See Fisher, *Statistical Methods*, Chap. IV, Sec. 21.03. (For full citations, see Sec. 7).

† See Sec. 6, below, for a discussion of published tables of this integral.

*Example* I.   The variance $\sigma_0^2$ of a particular dimension of a standard product is 3.2 cm.[2]   From measurements of the dimension on 10 items of a new product it is found that $s_1^2 = 4.8$, with 9 degrees of freedom.   Does the variance $\sigma_1^2$ of the dimension with the new product exceed the variance $\sigma_0^2$ of the dimension with the standard product?

If we conduct the test at the 0.05 level of significance, we find $s_1^2/\sigma_0^2 = 1.5 < 1.880 = \chi_{0.05}^2(9)/9$ and are led to conclude $\sigma_1^2/\sigma_0^2 \leq 1$; that is, we conclude that the variability of the new product does not exceed the standard in respect of the dimension in question.

## 2.4. Operating Characteristics of the Standard Procedure

The question arises, what are the chances of the standard rule leading to a wrong conclusion?   That is, when in fact $\sigma_1^2/\sigma_0^2 \leq 1$, how often will it lead to the conclusion $\sigma_1^2/\sigma_0^2 > 1$, termed an error of the first kind;[*] and when in fact $\sigma_1^2/\sigma_0^2 > 1$, how often will it lead to the conclusion $\sigma_1^2/\sigma_0^2 \leq 1$, termed an error of the second kind?   In other words, what are the operating characteristics of the procedure?

Clearly, the chance of an erroneous decision will depend on $\sigma_0^2$, $n_1$, and $\alpha$, as these quantities are involved explicitly in inequalities (1) and (2).   Intuition tells us that the chance of error depends also on the actual value of $\sigma_1^2/\sigma_0^2$, that is, upon

$$\rho = \frac{\sigma_1^2}{\sigma_0^2} \tag{4}$$

To begin with, it is clear from (3) that the probability of an error of the first kind is exactly $\alpha$ when $\rho = 1$.   Second, it will be shown in Sec. 6 that the probability of an error of the second kind will be exactly $\beta$ if $\rho$ has the value

$$\rho(\alpha,\beta,n_1) = \frac{\chi_\alpha^2(n_1)}{\chi_{1-\beta}^2(n_1)} \tag{5}$$

This equation has been used to compute tables of $\rho(\alpha,\beta,n_1)$ corresponding to $\alpha = 0.05$ and 0.01 and to a wide selection of values of $\beta$ and $n_1$.   These are published as Tables 8.1 and 8.2, pp. 272–273 and 276–277; the method of their construction and the accuracy of interpolation is discussed in Sec. 6.

The following examples illustrate the use of Tables 8.1 and 8.2 in the interpretation of the results of experiments already carried out and in the planning of experiments:

*Example* II.   For what value of the ratio $\sigma_1^2/\sigma_0^2$ is the probability of an error of the second kind equal to 0.50 when the test of Example I is employed?

[*] This terminology was introduced by Neyman and Pearson.

TABLE 8.1

VALUES OF $\rho(0.05,\beta,n_1)$ CORRESPONDING TO SELECTED VALUES OF $\beta$ AND $n_1$

| Degrees of freedom / $\beta$ | 0.005 | 0.01 | 0.025 | 0.05 | 0.10 | 0.25 | 0.50 | 0.75 | 0.90 | 0.95 | 0.975 | 0.99 | 0.995 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 97821. | 24454. | 3911.6 | 976.94 | 243.27 | 37.835 | 8.4439 | 2.9029 | 1.4198 | 1.0000 | 0.7646 | 0.5790 | 0.4875 |
| 2 | 597.65 | 298.07 | 118.33 | 58.404 | 28.433 | 10.413 | 4.3219 | 2.1610 | 1.3010 | 1.0000 | 0.8121 | 0.6505 | 0.5654 |
| 3 | 108.96 | 68.054 | 36.214 | 22.211 | 13.373 | 6.4450 | 3.3030 | 1.9022 | 1.2501 | 1.0000 | 0.8359 | 0.6888 | 0.6087 |
| 4 | 45.837 | 31.933 | 19.586 | 13.349 | 8.9202 | 4.9350 | 2.8265 | 1.7618 | 1.2196 | 1.0000 | 0.8514 | 0.7146 | 0.6385 |
| 5 | 26.887 | 19.972 | 13.318 | 9.6645 | 6.8748 | 4.1391 | 2.5441 | 1.6708 | 1.1986 | 1.0000 | 0.8627 | 0.7338 | 0.6609 |
| 6 | 18.634 | 14.438 | 10.176 | 7.6994 | 5.7127 | 3.6449 | 2.3544 | 1.6059 | 1.1829 | 1.0000 | 0.8714 | 0.7490 | 0.6789 |
| 7 | 14.220 | 11.353 | 8.3244 | 6.4905 | 4.9653 | 3.3061 | 2.2168 | 1.5566 | 1.1706 | 1.0000 | 0.8785 | 0.7614 | 0.6937 |
| 8 | 11.535 | 9.4184 | 7.1143 | 5.6748 | 4.4439 | 3.0583 | 2.1115 | 1.5175 | 1.1606 | 1.0000 | 0.8844 | 0.7719 | 0.7063 |
| 9 | 9.7520 | 8.1033 | 6.2654 | 5.0883 | 4.0591 | 2.8682 | 2.0280 | 1.4856 | 1.1522 | 1.0000 | 0.8894 | 0.7809 | 0.7172 |
| 10 | 8.4918 | 7.1562 | 5.6382 | 4.6461 | 3.7629 | 2.7173 | 1.9597 | 1.4589 | 1.1451 | 1.0000 | 0.8938 | 0.7888 | 0.7268 |
| 11 | 7.5580 | 6.4435 | 5.1563 | 4.3007 | 3.5274 | 2.5942 | 1.9026 | 1.4361 | 1.1389 | 1.0000 | 0.8976 | 0.7958 | 0.7353 |
| 12 | 6.8404 | 5.8887 | 4.7745 | 4.0233 | 3.3355 | 2.4917 | 1.8541 | 1.4163 | 1.1335 | 1.0000 | 0.9010 | 0.8020 | 0.7430 |
| 13 | 6.2726 | 5.4450 | 4.4646 | 3.7954 | 3.1758 | 2.4048 | 1.8122 | 1.3990 | 1.1287 | 1.0000 | 0.9040 | 0.8076 | 0.7499 |
| 14 | 5.8127 | 5.0821 | 4.2078 | 3.6046 | 3.0406 | 2.3300 | 1.7756 | 1.3837 | 1.1244 | 1.0000 | 0.9068 | 0.8128 | 0.7562 |
| 15 | 5.4328 | 4.7799 | 3.9916 | 3.4425 | 2.9246 | 2.2648 | 1.7432 | 1.3700 | 1.1205 | 1.0000 | 0.9093 | 0.8174 | 0.7620 |
| 16 | 5.1138 | 4.5243 | 3.8068 | 3.3029 | 2.8238 | 2.2075 | 1.7144 | 1.3577 | 1.1170 | 1.0000 | 0.9116 | 0.8218 | 0.7674 |
| 17 | 4.8422 | 4.3053 | 3.6471 | 3.1813 | 2.7354 | 2.1566 | 1.6885 | 1.3465 | 1.1138 | 1.0000 | 0.9138 | 0.8257 | 0.7723 |
| 18 | 4.6082 | 4.1154 | 3.5075 | 3.0743 | 2.6571 | 2.1111 | 1.6651 | 1.3362 | 1.1108 | 1.0000 | 0.9157 | 0.8295 | 0.7770 |
| 19 | 4.4044 | 3.9492 | 3.3844 | 2.9795 | 2.5872 | 2.0700 | 1.6438 | 1.3269 | 1.1081 | 1.0000 | 0.9175 | 0.8329 | 0.7813 |
| 20 | 4.2253 | 3.8025 | 3.2750 | 2.8948 | 2.5244 | 2.0328 | 1.6243 | 1.3182 | 1.1055 | 1.0000 | 0.9192 | 0.8361 | 0.7853 |
| 21 | 4.0667 | 3.6720 | 3.1772 | 2.8185 | 2.4676 | 1.9989 | 1.6064 | 1.3102 | 1.1032 | 1.0000 | 0.9208 | 0.8392 | 0.7891 |
| 22 | 3.9252 | 3.5551 | 3.0890 | 2.7496 | 2.4160 | 1.9678 | 1.5899 | 1.3028 | 1.1010 | 1.0000 | 0.9223 | 0.8420 | 0.7927 |
| 23 | 3.7982 | 3.4497 | 3.0092 | 2.6869 | 2.3689 | 1.9392 | 1.5746 | 1.2959 | 1.0989 | 1.0000 | 0.9238 | 0.8447 | 0.7961 |

TABLE 8.1    COMPARING TWO STANDARD DEVIATIONS    273

TABLE 8.1.—*(Continued)*

| De-grees of freedom / β | 0.005 | 0.01 | 0.025 | 0.05 | 0.10 | 0.25 | 0.50 | 0.75 | 0.90 | 0.95 | 0.975 | 0.99 | 0.995 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 24 | 3.6834 | 3.3543 | 2.9364 | 2.6296 | 2.3256 | 1.9128 | 1.5604 | 1.2894 | 1.0970 | 1.0000 | 0.9251 | 0.8473 | 0.7993 |
| 25 | 3.5792 | 3.2673 | 2.8699 | 2.5769 | 2.2857 | 1.8884 | 1.5472 | 1.2834 | 1.0951 | 1.0000 | 0.9263 | 0.8497 | 0.8023 |
| 26 | 3.4842 | 3.1878 | 2.8088 | 2.5284 | 2.2488 | 1.8656 | 1.5348 | 1.2777 | 1.0934 | 1.0000 | 0.9275 | 0.8520 | 0.8052 |
| 27 | 3.3972 | 3.1147 | 2.7525 | 2.4836 | 2.2145 | 1.8443 | 1.5231 | 1.2723 | 1.0918 | 1.0000 | 0.9287 | 0.8541 | 0.8080 |
| 28 | 3.3172 | 3.0474 | 2.7004 | 2.4420 | 2.1826 | 1.8245 | 1.5122 | 1.2672 | 1.0902 | 1.0000 | 0.9297 | 0.8562 | 0.8106 |
| 29 | 3.2434 | 2.9851 | 2.6520 | 2.4032 | 2.1529 | 1.8058 | 1.5019 | 1.2624 | 1.0888 | 1.0000 | 0.9308 | 0.8582 | 0.8132 |
| 30 | 3.1750 | 2.9273 | 2.6070 | 2.3670 | 2.1250 | 1.7883 | 1.4921 | 1.2578 | 1.0874 | 1.0000 | 0.9318 | 0.8601 | 0.8156 |
| 40 | 2.6928 | 2.5157 | 2.2821 | 2.1034 | 1.9194 | 1.6565 | 1.4175 | 1.2223 | 1.0763 | 1.0000 | 0.9396 | 0.8755 | 0.8351 |
| 50 | 2.4117 | 2.2724 | 2.0862 | 1.9418 | 1.7911 | 1.5720 | 1.3683 | 1.1983 | 1.0687 | 1.0000 | 0.9452 | 0.8864 | 0.8492 |
| 60 | 2.2255 | 2.1097 | 1.9535 | 1.8311 | 1.7022 | 1.5123 | 1.3328 | 1.1807 | 1.0630 | 1.0000 | 0.9494 | 0.8948 | 0.8600 |
| 70 | 2.0920 | 1.9922 | 1.8568 | 1.7498 | 1.6362 | 1.4673 | 1.3057 | 1.1670 | 1.0585 | 1.0000 | 0.9527 | 0.9015 | 0.8687 |
| 80 | 1.9909 | 1.9029 | 1.7826 | 1.6870 | 1.5850 | 1.4320 | 1.2842 | 1.1560 | 1.0549 | 1.0000 | 0.9555 | 0.9070 | 0.8758 |
| 90 | 1.9114 | 1.8322 | 1.7235 | 1.6368 | 1.5438 | 1.4034 | 1.2665 | 1.1469 | 1.0519 | 1.0000 | 0.9578 | 0.9116 | 0.8819 |
| 100 | 1.8468 | 1.7747 | 1.6753 | 1.5956 | 1.5098 | 1.3795 | 1.2518 | 1.1393 | 1.0493 | 1.0000 | 0.9597 | 0.9156 | 0.8871 |
| 120 | 1.7479 | 1.6861 | 1.6005 | 1.5315 | 1.4566 | 1.3420 | 1.2282 | 1.1270 | 1.0452 | 1.0000 | 0.9629 | 0.9221 | 0.8957 |
| ∞ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |

Explanation

Let (a) $s_1^2$ be an estimate of $\sigma_1^2$ such that $\dfrac{n_1 s_1^2}{\sigma_1^2}$ is distributed as $\chi^2$ for $n_1$ degrees of freedom

(b) $\chi_\beta^2(n_1)$ be the value of $\chi^2$ for $n_1$ degrees of freedom exceeded with probability $\epsilon$

(c) the decision $\dfrac{\sigma_1^2}{\sigma_0^2} > 1$, or the decision $\dfrac{\sigma_1^2}{\sigma_0^2} \leq 1$, be made according as $\dfrac{s_1^2}{\sigma_0^2}$ exceeds or does not exceed $\dfrac{1}{n_1}\chi_{0.05}^2(n_1)$

Then the probability is $\beta$ that the decision $\dfrac{\sigma_1^2}{\sigma_0^2} \leq 1$ will be made when $\dfrac{\sigma_1^2}{\sigma_0^2} = \rho(0.05, \beta, n_1) = \dfrac{\chi_{0.05}^2(n_1)}{\chi_{1-\beta}^2(n_1)}$

For $\alpha = 0.05$, $n_1 = 9$, and $\beta = 0.50$, we find from Table 8.1 that $\rho(0.05,0.50,9) = 2.0280$. This means that if $\sigma_1^2$ is roughly double $\sigma_0^2$ (that is, if the standard deviation of the new product is roughly 1.4 times that of the standard) and the test is conducted at the 0.05 level of significance with $s_1^2$ based on only 9 degrees of

**.05 Level of Significance**
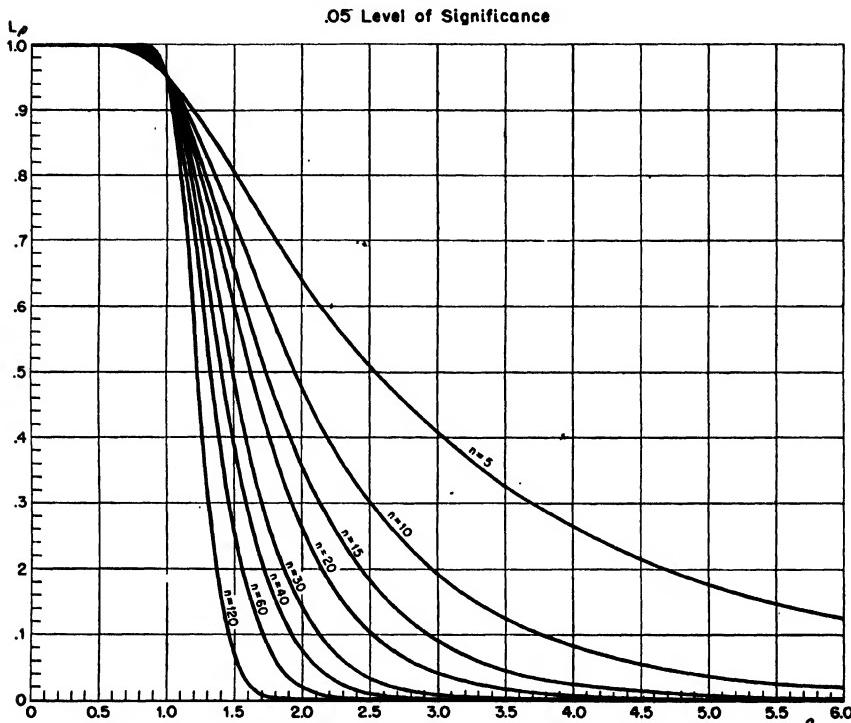


FIG. 8.1.—Operating-characteristic curves for experiments to determine whether a new product (process) is more variable than a standard product (process) of known variability, 0.05 level of significance.

freedom, then there is approximately an even chance that the test will lead to the wrong conclusion, $\sigma_1^2/\sigma_0^2 \leq 1$. If, to ensure a smaller risk of committing an error of the first kind, the test is conducted at the 0.01 level of significance, the situation is worse, for

$$\rho(0.01,0.50,9) = 2.5970$$

(from Table 8.2). Clearly 9 degrees of freedom for $s_1^2$ are too few if a decision that $\sigma_1^2/\sigma_0^2 \leq 1$ is a serious error from the practical viewpoint when $\sigma_1$ is double $\sigma_0$.

*Example* III. If a decision that $\sigma_1^2/\sigma_0^2 \leq 1$ is a serious error from the practical viewpoint when $\sigma_1 = 1.5\sigma_0$ ($\rho = 2.25$), and it is

desired to keep the risk of such an error below 0.05 ($\beta = 0.05$) when the test is conducted at the 0.05 level of significance ($\alpha = 0.05$), how many degrees of freedom will be needed for $s_1^2$?

On looking in the $\beta = 0.05$ column of Table 8.1, we find that $n_1 = 30$ is not quite sufficient, since $\rho(0.05,0.05,30) = 2.3670$; but



.01 Level of Significance
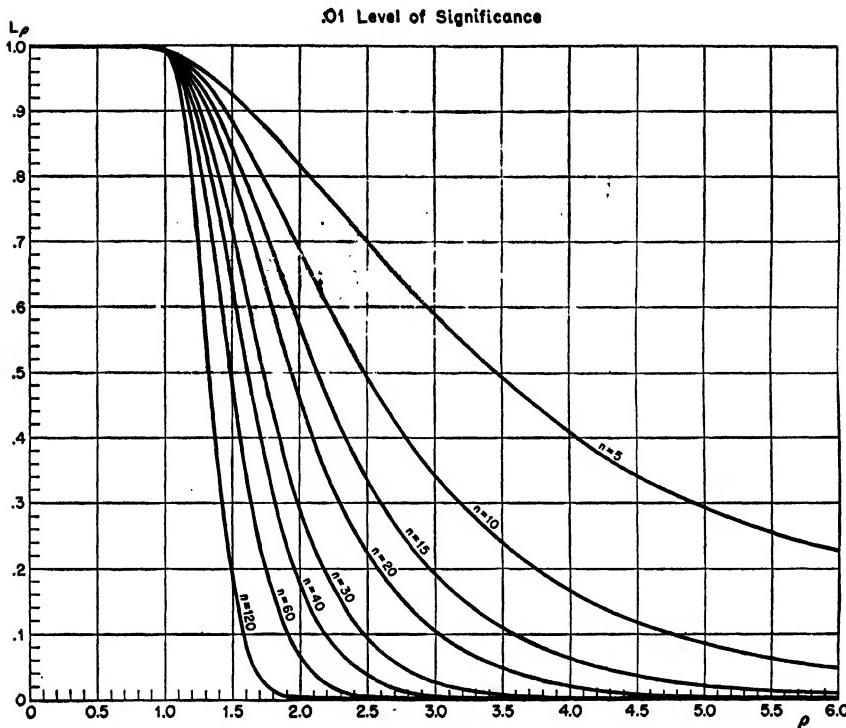
Fig. 8.2.—Operating-characteristic curves for experiments to determine whether a new product (process) is more variable than a standard product (process) of known variability, 0.01 level of significance.

$n_1 = 40$ is ample, since $\rho(0.05,0.05,40) = 2.1034$. By interpolation, using $1/\sqrt{n_1}$ as argument, it is found that 34 (33.7) degrees of freedom are needed (linear interpolation gives 34.4, which would be taken as 35).

If $\sigma_0 = 1.79$ cm. ($\sigma_0^2 = 3.2$ cm.²) as in Example I, so that $\sigma_1 = 1.5\sigma_0 = 2.69$ cm., and it is felt that this is too large an excess to have undetected 1 time in 20, but failure to detect that $\sigma_1 > \sigma_0$ when $\sigma_1 = 2.00$ cm., say, could be tolerated 1 time in 20, then $\rho = 1.25$, and it is clear from an examination of Table 8.1 that a very large number of degrees of freedom, over 400, are needed—

TABLE 8.2

VALUES OF $\rho(0.01,\beta,n_1)$ CORRESPONDING TO SELECTED VALUES OF $\beta$ AND $n_1$

| Degrees of freedom \ $\beta$ | 0.005 | 0.01 | 0.025 | 0.05 | 0.10 | 0.25 | 0.50 | 0.75 | 0.90 | 0.95 | 0.975 | 0.99 | 0.995 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 168950. | 42237. | 6756.0 | 1687.4 | 420.18 | 65.349 | 14.584 | 5.0139 | 2.4523 | 1.7272 | 1.3207 | 1.0000 | 0.8421 |
| 2 | 918.73 | 458.21 | 181.89 | 89.781 | 43.709 | 16.008 | 6.6439 | 3.3219 | 2.0000 | 1.5372 | 1.2484 | 1.0000 | 0.8692 |
| 3 | 158.18 | 98.796 | 52.573 | 32.244 | 19.414 | 9.3564 | 4.7950 | 2.7614 | 1.8148 | 1.4517 | 1.2136 | 1.0000 | 0.8837 |
| 4 | 64.142 | 44.686 | 27.408 | 18.681 | 12.482 | 6.9058 | 3.9553 | 2.4654 | 1.7066 | 1.3994 | 1.1915 | 1.0000 | 0.8934 |
| 5 | 36.640 | 27.217 | 18.150 | 13.170 | 9.3686 | 5.6406 | 3.4670 | 2.2769 | 1.6334 | 1.3627 | 1.1756 | 1.0000 | 0.9007 |
| 6 | 24.880 | 19.278 | 13.587 | 10.280 | 7.6275 | 4.8665 | 3.1435 | 2.1442 | 1.5794 | 1.3352 | 1.1635 | 1.0000 | 0.9064 |
| 7 | 18.676 | 14.911 | 10.933 | 8.5244 | 6.5212 | 4.3422 | 2.9114 | 2.0444 | 1.5374 | 1.3134 | 1.1538 | 1.0000 | 0.9111 |
| 8 | 14.943 | 12.202 | 9.2168 | 7.3519 | 5.7573 | 3.9621 | 2.7355 | 1.9660 | 1.5036 | 1.2955 | 1.1457 | 1.0000 | 0.9151 |
| 9 | 12.488 | 10.377 | 8.0233 | 6.5159 | 5.1980 | 3.6729 | 2.5970 | 1.9024 | 1.4755 | 1.2806 | 1.1389 | 1.0000 | 0.9185 |
| 10 | 10.766 | 9.0725 | 7.1480 | 5.8902 | 4.7705 | 3.4449 | 2.4845 | 1.8495 | 1.4518 | 1.2678 | 1.1331 | 1.0000 | 0.9214 |
| 11 | 9.4979 | 8.0973 | 6.4797 | 5.4046 | 4.4328 | 3.2601 | 2.3910 | 1.8047 | 1.4313 | 1.2567 | 1.1280 | 1.0000 | 0.9241 |
| 12 | 8.5291 | 7.3425 | 5.9533 | 5.0166 | 4.1589 | 3.1069 | 2.3118 | 1.7660 | 1.4134 | 1.2469 | 1.1234 | 1.0000 | 0.9264 |
| 13 | 7.7666 | 6.7419 | 5.5280 | 4.6994 | 3.9322 | 2.9775 | 2.2438 | 1.7323 | 1.3976 | 1.2382 | 1.1194 | 1.0000 | 0.9285 |
| 14 | 7.1518 | 6.2529 | 5.1773 | 4.4351 | 3.7411 | 2.8667 | 2.1846 | 1.7025 | 1.3835 | 1.2304 | 1.1157 | 1.0000 | 0.9305 |
| 15 | 6.6460 | 5.8474 | 4.8830 | 4.2113 | 3.5777 | 2.7706 | 2.1325 | 1.6760 | 1.3708 | 1.2233 | 1.1124 | 1.0000 | 0.9322 |
| 16 | 6.2229 | 5.5056 | 4.6325 | 4.0193 | 3.4363 | 2.6863 | 2.0862 | 1.6521 | 1.3593 | 1.2169 | 1.1094 | 1.0000 | 0.9338 |
| 17 | 5.8640 | 5.2138 | 4.4167 | 3.8526 | 3.3126 | 2.6117 | 2.0448 | 1.6306 | 1.3488 | 1.2110 | 1.1066 | 1.0000 | 0.9353 |
| 18 | 5.5557 | 4.9616 | 4.2287 | 3.7065 | 3.2035 | 2.5451 | 2.0075 | 1.6110 | 1.3392 | 1.2056 | 1.1040 | 1.0000 | 0.9367 |
| 19 | 5.2880 | 4.7415 | 4.0634 | 3.5772 | 3.1063 | 2.4853 | 1.9736 | 1.5931 | 1.3304 | 1.2006 | 1.1016 | 1.0000 | 0.9380 |
| 20 | 5.0534 | 4.5477 | 3.9169 | 3.4621 | 3.0192 | 2.4312 | 1.9427 | 1.5766 | 1.3222 | 1.1960 | 1.0994 | 1.0000 | 0.9392 |
| 21 | 4.8461 | 4.3758 | 3.7861 | 3.3587 | 2.9406 | 2.3820 | 1.9143 | 1.5614 | 1.3146 | 1.1917 | 1.0973 | 1.0000 | 0.9404 |
| 22 | 4.6617 | 4.2221 | 3.6686 | 3.2655 | 2.8693 | 2.3370 | 1.8882 | 1.5473 | 1.3075 | 1.1876 | 1.0954 | 1.0000 | 0.9414 |
| 23 | 4.4964 | 4.0839 | 3.5623 | 3.1808 | 2.8043 | 2.2957 | 1.8641 | 1.5341 | 1.3009 | 1.1838 | 1.0936 | 1.0000 | 0.9424 |
| 24 | 4.3474 | 3.9589 | 3.4658 | 3.1036 | 2.7448 | 2.2577 | 1.8417 | 1.5219 | 1.2947 | 1.1803 | 1.0919 | 1.0000 | 0.9434 |

TABLE 8.2    COMPARING TWO STANDARD DEVIATIONS    277

TABLE 8.2.—(Continued)

| Degrees of freedom \ β | 0.005 | 0.01 | 0.025 | 0.05 | 0.10 | 0.25 | 0.50 | 0.75 | 0.90 | 0.95 | 0.975 | 0.99 | 0.995 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 25 | 4.2125 | 3.8454 | 3.3777 | 3.0328 | 2.6900 | 2.2225 | 1.8209 | 1.5104 | 1.2889 | 1.1769 | 1.0902 | 1.0000 | 0.9443 |
| 26 | 4.0896 | 3.7417 | 3.2969 | 2.9678 | 2.6395 | 2.1897 | 1.8014 | 1.4997 | 1.2834 | 1.1738 | 1.0887 | 1.0000 | 0.9452 |
| 27 | 3.9774 | 3.6466 | 3.2225 | 2.9077 | 2.5927 | 2.1593 | 1.7832 | 1.4895 | 1.2782 | 1.1708 | 1.0872 | 1.0000 | 0.9460 |
| 28 | 3.8743 | 3.5591 | 3.1538 | 2.8520 | 2.5491 | 2.1308 | 1.7661 | 1.4800 | 1.2733 | 1.1679 | 1.0859 | 1.0000 | 0.9468 |
| 29 | 3.7792 | 3.4783 | 3.0901 | 2.8003 | 2.5085 | 2.1042 | 1.7500 | 1.4710 | 1.2686 | 1.1652 | 1.0845 | 1.0000 | 0.9475 |
| 30 | 3.6914 | 3.4034 | 3.0310 | 2.7520 | 2.4706 | 2.0791 | 1.7348 | 1.4624 | 1.2642 | 1.1626 | 1.0833 | 1.0000 | 0.9482 |
| 40 | 3.0759 | 2.8736 | 2.6067 | 2.4026 | 2.1924 | 1.8922 | 1.6021 | 1.3962 | 1.2294 | 1.1423 | 1.0733 | 1.0000 | 0.9539 |
| 50 | 2.7207 | 2.5635 | 2.3535 | 2.1906 | 2.0206 | 1.7734 | 1.5356 | 1.3518 | 1.2056 | 1.1281 | 1.0663 | 1.0000 | 0.9580 |
| 60 | 2.4871 | 2.3577 | 2.1832 | 2.0464 | 1.9023 | 1.6901 | 1.4895 | 1.3195 | 1.1879 | 1.1176 | 1.0610 | 1.0000 | 0.9612 |
| 70 | 2.3206 | 2.2100 | 2.0597 | 1.9410 | 1.8151 | 1.6277 | 1.4484 | 1.2945 | 1.1742 | 1.1093 | 1.0568 | 1.0000 | 0.9636 |
| 80 | 2.1951 | 2.0980 | 1.9654 | 1.8600 | 1.7476 | 1.5789 | 1.4159 | 1.2746 | 1.1631 | 1.1026 | 1.0535 | 1.0000 | 0.9657 |
| 90 | 2.0967 | 2.0098 | 1.8907 | 1.7955 | 1.6935 | 1.5394 | 1.3893 | 1.2581 | 1.1539 | 1.0970 | 1.0506 | 1.0000 | 0.9674 |
| 100 | 2.0171 | 1.9383 | 1.8297 | 1.7427 | 1.6490 | 1.5067 | 1.3672 | 1.2443 | 1.1461 | 1.0922 | 1.0482 | 1.0000 | 0.9689 |
| 120 | 1.8956 | 1.8286 | 1.7358 | 1.6609 | 1.5797 | 1.4553 | 1.3320 | 1.2222 | 1.1335 | 1.0845 | 1.0443 | 1.0000 | 0.9713 |
| ∞ | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |

Explanation

Let (a) $s_1^2$ be an estimate of $\sigma_1^2$ such that $\dfrac{n_1 s_1^2}{\sigma_1^2}$ is distributed as $\chi^2$ for $n_1$ degrees of freedom

(b) $\chi^2(n_1)$ be the value of $\chi^2$ for $n_1$ degrees of freedom exceeded with probability $\epsilon$

(c) the decision $\dfrac{\sigma_1^2}{\sigma_0^2} > 1$, or the decision $\dfrac{\sigma_1^2}{\sigma_0^2} \leq 1$, be made according as $\dfrac{s_1^2}{\sigma_0^2}$ exceeds or does not exceed $\dfrac{1}{n_1}\chi_{0.01}^2(n_1)$

Then the probability is $\beta$ that the decision $\dfrac{\sigma_1^2}{\sigma_0^2} \leq 1$ will be made when $\dfrac{\sigma_1^2}{\sigma_0^2} = \rho(0.01, \beta, n_1) = \dfrac{\chi_{0.01}^2(n_1)}{\chi_{1-\beta}^2(n_1)}$.

an astonishingly large number to persons who have never investigated this matter before.*

As is pointed out in the footnote to the tables, $\rho(\alpha,\beta,n_1)$ is the value of $\sigma_1^2/\sigma_0^2$ for which the probability of the decision $\sigma_1^2/\sigma_0^2 \leq 1$ equals $\beta$ when the test is conducted at the $\alpha$ level of significance with $n_1$ degrees of freedom. When $\rho(\alpha,\beta,n_1) \leq 1$, $\beta$ can no longer be interpreted as the probability of an error of the second kind, since the decision $\sigma_1^2/\sigma_0^2 \leq 1$ is correct under these circumstances.† The values of $\rho(\alpha,\beta,n_1) < 1$ have been included in the tables since they are useful for drawing the operating-characteristic curves.

Let $L_\rho = L(\alpha,n_1,\rho)$ denote the probability that the decision $\sigma_1^2/\sigma_0^2 \leq 1$ will be made when $\sigma_1^2/\sigma_0^2 = \rho$ and the test defined by formulas (1) and (2) is conducted at the $\alpha$ level of significance with $n_1$ degrees of freedom for $s_1^2$. Then $L_\rho$ plotted as a function of $\rho$, for fixed values of $\alpha$ and $n_1$, is termed the operating-characteristic curve of the test so defined. Since the probability that the decision $\sigma_1^2/\sigma_0^2 > 1$ will be reached is $1 - L_\rho$, in view of the exhaustive character of inequalities (1) and (2), it is clear that the curve $L_\rho$ completely specifies the operating characteristics of the test, that is, the expected performance of the test in repeated applications. Thus, it is not necessary to plot the complementary curve $1 - L_\rho$, except perhaps for psychological reasons. A probability of 0.90 that a correct decision will be made may evoke less concern about the stringency of the test than a statement that the probability of an incorrect decision is 0.10.

Figure 8.1 shows the operating-characteristic curves corresponding to 5, 10, 15, 20, 30, 40, 60, and 120 degrees of freedom when $\alpha = 0.05$; Fig. 8.2 shows the corresponding curves when $\alpha = 0.01$.

### 3. EXPERIMENTS TO DETERMINE WHETHER A NEW PRODUCT IS LESS VARIABLE THAN A STANDARD PRODUCT

### 3.1. Problem

There are not many situations in which a new product is preferred to a standard product only if its variability exceeds that of the standard. Consequently, a test of the hypothesis $\sigma_1^2/\sigma_0^2 \geq 1$ relative to the alterna-

---

* The following formula may be used when $n$ is large, say more than 30:

$$n = \frac{1}{2}\left(\frac{K_\alpha \sigma_0 + K_\beta \sigma_1}{\sigma_1 - \sigma_0}\right)^2$$

$$= \frac{1}{2}\left(\frac{K_\alpha + K_\beta \sqrt{\rho}}{\sqrt{\rho} - 1}\right)^2$$

where $K_\epsilon$ denotes a normal deviate exceeded with probability $\epsilon$ and may be found from Table 1.1 of Chap. 1.

† From (5) it is clear that $\rho(\alpha,\beta,n_1) = 1$ for $\beta = 1 - \alpha$, and is less than 1 for $\beta > 1 - \alpha$, since $\chi_\epsilon^2$ increases as $\epsilon$ decreases, whatever the value of $n$.

tive $\sigma_1^2/\sigma_0^2 < 1$ is rarely employed. Nevertheless, there are situations in which too small a variability is disadvantageous. For instance, an average gunner is more likely to bring down a bird with a shotgun than with a rifle. It would not be advisable to reduce the dispersion of the shot to nearly zero.

In such cases the standard procedure is modified to

(a) Conclude $\sigma_1^2/\sigma_0^2 \geq 1$ whenever

$$\frac{s_1^2}{\sigma_0^2} \geq \frac{\chi_{1-\alpha}^2(n_1)}{n_1} \tag{6}$$

(b) Conclude $\sigma_1^2/\sigma_0^2 < 1$ whenever

$$\frac{s_1^2}{\sigma_0^2} < \frac{\chi_{1-\alpha}^2(n_1)}{n_1} \tag{7}$$

In connection with this rule, the decision $\sigma_1^2/\sigma_0^2 < 1$ when in fact $\sigma_1^2/\sigma_0^2 \geq 1$ is an error of the first kind, whence the maximum value of the probability of an error of the first kind is $\alpha$, which is the probability of the decision $\sigma_1^2/\sigma_0^2 < 1$ when $\sigma_1^2/\sigma_0^2 = 1$. Accordingly, the statistical test defined by (6) and (7) is said to be conducted at the $\alpha$ level of significance, although $\chi_{1-\alpha}^2(n_1)$ is the $100(1 - \alpha)$ percentage point of $\chi^2$ for $n_1$ degrees of freedom.

The value of $\rho' = \sigma_1^2/\sigma_0^2$ for which the probability of a decision $\sigma_1^2/\sigma_0^2 \geq 1$ equals $\beta$ when the test specified by inequalities (6) and (7) is employed is readily shown to be

$$\rho'(\alpha,\beta,n_1) = \frac{\chi_{1-\alpha}^2(n_1)}{\chi_\beta^2(n_1)} \tag{8}$$

$$= \frac{1}{\rho(\beta,\alpha,n_1)}$$

Consequently, values of $\rho'$ are found by interchanging $\alpha$ and $\beta$ and taking the reciprocal of the corresponding $\rho$. Thus

$$\rho'(0.05,0.05,n_1) = \frac{1}{\rho(0.05,0.05,n_1)} \tag{9}$$

and

$$\rho'(0.01,0.05,n_1) = \frac{1}{\rho(0.05,0.01,n_1)} \tag{10}$$

can be obtained from Table 8.1 by taking reciprocals of the relevant entries; and values of

$$\rho'(0.01,0.01,n_1) = \frac{1}{\rho(0.01,0.01,n_1)} \tag{11}$$

and

$$\rho'(0.05,0.01,n_1) = \frac{1}{\rho(0.01,0.05,n_1)} \tag{12}$$

can be obtained in similar manner from Table 8.2.

### 3.2. Illustration

*Example* IV. The scores of students taking a standard examination have a standard deviation of 20 points ($\sigma_0 = 20$). An untried examination will be acceptable if with it students' scores have a standard deviation not less than 20 points ($\sigma_1^2/\sigma_0^2 \geq 1$), and the authorities are willing to accept a risk of 0.05 ($\alpha = 0.05$) of erroneously rejecting an examination for which $\sigma_1 = \sigma_0 = 20$. On the other hand, acceptance of an examination for which $\sigma_1 = 10$ ($\rho' = \frac{1}{4}$) would be a serious error, and it is desired to hold the risk of such an occurrence below 0.01. How many students will be needed in a preliminary tryout of the examination?

For $\alpha = 0.05$ and $\beta = 0.01$, we seek the value of $n_1$ for which $\rho'(0.05,0.01,n_1)$ is 0.25, that is, the value $n_1$ for which $\rho(0.01,0.05,n_1)$ is 4. In Table 8.2 we see that

$$\rho(0.01,0.05,16) = 4.0193 \tag{13}$$
$$\rho(0.01,0.05,17) = 3.8526 \tag{14}$$

Accordingly, $n_1 \geq 17$ will be required, which means that the smallest number of students ($n_1 + 1$) that will suffice is 18.

### 4. EXPERIMENTS TO DETERMINE WHETHER A PARTICULAR ONE OF TWO NEW PRODUCTS IS MORE VARIABLE THAN THE OTHER

### 4.1. Problem

When the variance of the measurable characteristic is unknown for both products, as it generally is when both are new, the rules of procedure given in Secs. 2 and 3 cannot be applied. Let $\sigma_1^2$ denote the true variance of the measurable characteristic in the case of product I, let $\sigma_2^2$ denote the variance in the case of product II, and let $s_1^2$ and $s_2^2$ denote estimates of $\sigma_1^2$ and $\sigma_2^2$ computed from measurements on I and II, respectively. A choice between the conclusion $\sigma_1^2/\sigma_2^2 \leq 1$ and the conclusion $\sigma_1^2/\sigma_2^2 > 1$ can then be based on the observed values of $s_1^2$ and $s_2^2$, for example, on the observed value of their ratio $s_1^2/s_2^2$. That is, it may be agreed in advance that either the conclusion $\sigma_1^2/\sigma_2^2 > 1$ or the conclusion $\sigma_1^2/\sigma_2^2 \leq 1$ will be accepted according as $s_1^2/s_2^2$ exceeds or does not exceed some number $C$ chosen in advance.

In selecting $C$ it is necessary to distinguish between the following two situations: (1) For various reasons (for example, cost, convenience, and so forth) product I is preferred a priori and should be chosen unless $\sigma_1^2$ is substantially larger than $\sigma_2^2$. (2) Product II is to be chosen if $\sigma_1^2 > \sigma_2^2$, but it is immaterial which is chosen when $\sigma_1^2 = \sigma_2^2$.

## 4.2. Standard Procedure When Product I Is to Be Chosen Unless $\sigma_1^2$ Is Larger Than $\sigma_2^2$

**4.2.1. Procedure.**—The standard procedure in case (1) is to compute from the data estimates $s_1^2$ and $s_2^2$ such that in repeated sampling $n_1 s_1^2/\sigma_1^2$ and $n_2 s_2^2/\sigma_2^2$ will be independently as $\chi^2$ for $n_1$ and $n_2$ degrees of freedom, respectively, and to base the choice on the following rule:

(a) Accept product I whenever

$$\frac{s_1^2}{s_2^2} \leq F_\alpha(n_1, n_2) \tag{15}$$

(b) Accept product II whenever

$$\frac{s_1^2}{s_2^2} > F_\alpha(n_1, n_2) \tag{16}$$

where $F_\alpha(n_1, n_2)$ is the $\alpha$ level of significance of the variance ratio distribution

$$\alpha = P\{F(n_1, n_2) > F_\alpha(n_1, n_2)\}$$

$$= \frac{1}{\Gamma\left(\dfrac{n_1}{2}\right)\Gamma\left(\dfrac{n_2}{2}\right)} \frac{\Gamma\left(\dfrac{n_1 + n_2}{2}\right)}{}\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} \int_{F_\alpha(n_1,n_2)}^{\infty} \frac{F^{\frac{n_1-2}{2}}\, dF}{\left(1 + \dfrac{n_1 F}{n_2}\right)^{\frac{n_1+n_2}{2}}} \tag{17}$$

Since, under the assumptions regarding the distribution of $s_1^2$ and $s_2^2$, the ratio $(s_1^2/\sigma_1^2)/(s_2^2/\sigma_2^2)$ is distributed as $F$ for $n_1$ and $n_2$ degrees of freedom whatever the values of $\sigma_1^2$ and $\sigma_2^2$, it is clear from (17) that the probability of (16) occurring when $\sigma_1^2 = \sigma_2^2$ is exactly $\alpha$.

Accordingly, if choosing product II when its variability is not less than that of product I (that is, when $\sigma_1^2/\sigma_2^2 \leq 1$) is termed an error of the first kind, clearly $\alpha$ is the maximum value of the probability of an error of the first kind when the standard procedure is followed and the distributions of $s_1^2$ and $s_2^2$ are as assumed.    The test is thus said to be conducted at the $\alpha$ level of significance.

Let the true value of the ratio $\sigma_1^2/\sigma_2^2$ be denoted by

$$\phi = \frac{\sigma_1^2}{\sigma_2^2} \tag{18}$$

Then it is of interest to determine the value of $\phi$ for which the probability is $\beta$ that product I will be chosen by application of (15) and (16).   The answer is readily found from the relation*

* Shortly after this formula was derived, a British document was received in which essentially the same formula is given and attributed to D. V. Lindley; see BAINES, *Economical Design.*

$$\phi(\alpha,\beta,n_1,n_2) = \frac{F_\alpha(n_1,n_2)}{F_{1-\beta}(n_1,n_2)} \tag{19}$$
$$= F_\alpha(n_1,n_2) \cdot F_\beta(n_2,n_1)$$

Extensive tables of $\phi(\alpha,\beta,n_1,n_2)$ corresponding to $\alpha = 0.05$ and $0.01$ and to a range of values of $\beta$, $n_1$, and $n_2$ are presented as Tables 8.3 and 8.4, pp. 284–295 and 298–309. The use of formula (19) in the computation of these tables is explained in Sec. 6, which also includes a discussion of methods of interpolation. Figures 8.3 and 8.4 show the operating-characteristic curves plotted from the tables, for $n_1 = n_2 = 5, 10, 15, 20, 30, 40, 60, 120$.

**4.2.2. Two Special Cases.**—Before illustrating the use and interpretation of Tables 8.3 and 8.4, we may notice two important special cases involving $n_1 = n_2$:

(a) When $n_1 = n_2 = n$ and $\beta = 0.50$

$$F_{0.50}(n,n) = 1 \tag{20}$$

for all values of $n$, whence

$$\phi(\alpha,0.50,n,n) = F_\alpha(n,n) \tag{21}$$

In other words, when the degrees of freedom for $s_1^2$ and $s_2^2$ are equal, the probability of an error of the second kind is $\frac{1}{2}$ when $\sigma_1^2/\sigma_2^2 = F_\alpha(n,n)$, the critical ratio employed in the test. Thus $F_\alpha(n,n)$ may be termed the indifference value of $\sigma_1^2/\sigma_2^2$, that is, the value of $\sigma_1^2/\sigma_2^2$ for which the test is equally likely to select either product.

(b) When $n_1 = n_2 = n$ and $\beta = \alpha$

$$\phi(\alpha,\alpha,n,n) = F_\alpha(n,n) \tag{22}$$

In other words, when the degrees of freedom for $s_1^2$ and $s_2^2$ are equal, the probability of an error of the second kind is $\alpha$ when $\sigma_1/\sigma_2$, the ratio of the true standard deviations, equals $F_\alpha(n,n)$, the critical ratio employed in the test. That is, the risk of an error of the second kind when $\sigma_1/\sigma_2 = F_\alpha(n,n)$ is the same as the risk of an error of the first kind when $\sigma_1/\sigma_2 = 1$.

### 4.2.3. Illustrations

*Example* V. A choice between two products is to be made, the relative variability of the two products being relevant but of secondary importance in the sense that it is immaterial which is chosen if $\sigma_1^2/\sigma_2^2 = 2$, product I being preferred if $\sigma_1^2/\sigma_2^2 < 2$, product II if $\sigma_1^2/\sigma_2^2 > 2$. If the choice is to be based on (15) and (16), using the 0.05 level of significance ($\alpha = 0.05$), and if $s_1^2$ and $s_2^2$ are to be based on an equal number ($n_1 = n_2 = n$) degrees of freedom, how many degrees of freedom will be needed for each, and for what value of

$\sigma_1^2/\sigma_2^2$ is the probability 0.95 that product II will be chosen ($\beta = 0.05$)?
From the first special case of Sec. 4.2.2, we have

$$\phi(0.05, 0.50, n, n) = F_{0.05}(n, n),$$

and from tables of the $F$ distribution we find this to be 1.9838 when $n = 24$. Thus, if 24 degrees of freedom are used for each product, the probability that product II will be chosen is 0.05 if $\sigma_1^2/\sigma_2^2 = 1$, 0.50 if $\sigma_1^2/\sigma_2^2 = 2$, and (according to case (b) of Sec 4.2.2) 0.95 if $\sigma_1^2/\sigma_2^2 = 4$.

*Example* VI. If only 30 degrees of freedom are available for $s_1^2$ and if a test is conducted at the 0.05 level of significance, how many degrees of freedom will be needed for $s_2^2$ in order to ensure that the probability is only 0.05 of product I being chosen when $\sigma_1/\sigma_2 = 1.5$?

The desired stringency cannot be achieved, since Table 8.3 shows $\phi(0.05, 0.05, 30, \infty) = 2.367$, which exceeds $\sigma_1^2/\sigma_2^2 = 2.25$.

## 4.3. Procedure When It Is Immaterial Which Product Is Chosen If $\sigma_1^2 = \sigma_2^2$

This situation arises when variability is the primary basis for choosing between two products. The following rule is appropriate to this situation:

(a) Accept product I whenever

$$\frac{s_1^2}{s_2^2} \le F_{0.50}(n_1, n_2) \tag{23}$$

(b) Accept product II whenever

$$\frac{s_1^2}{s_2^2} > F_{0.50}(n_1, n_2) \tag{24}$$

where $F_{0.50}(n_1, n_2)$ is the median of the $F$ distribution for $n_1$ and $n_2$ degrees of freedom. Note that $F_{0.50}(n_1, n_2) \ne 1$ unless $n_1 = n_2$.

By setting $\alpha = 0.50$ in (19), it is readily seen that when the choice is based on (23) and (24) the probability is $\beta$ that product I will be chosen if

$$\frac{\sigma_1^2}{\sigma_2^2} = \phi(0.50, \beta, n_1, n_2) = \frac{F_{0.50}(n_1, n_2)}{F_{1-\beta}(n_1, n_2)} \tag{25}$$

Conversely, the probability is $\beta$ that product II will be chosen when

$$\frac{\sigma_1^2}{\sigma_2^2} = \phi(0.50, 1 - \beta, n_1, n_2) = \frac{F_{0.50}(n_1, n_2)}{F_\beta(n_1, n_2)} \tag{26}$$

## TABLE 8.3

### VALUES OF $\phi(0.05,\beta,n_1,n_2)$ CORRESPONDING TO SELECTED VALUES OF $\beta$, $n_1$, AND $n_2$

$$\beta = 0.005$$

| $n_2$ \ $n_1$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 15 | 20 | 24 | 30 | 40 | 60 | 120 | ∞ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2617000. | 39600. | 11980. | 7037. | 5244. | 4360. | 3844. | 3509. | 3275. | 3102. | 2887. | 2656. | 2466. | 2379. | 2296. | 2217. | 2142. | 2071. | 2004. |
| 2 | 370300. | 3781. | 954.3 | 505.9 | 353.4 | 281.1 | 240.1 | 213.9 | 195.9 | 182.8 | 165.2 | 149.6 | 135.9 | 129.6 | 123.7 | 118.1 | 112.9 | 107.9 | 108.3 |
| 3 | 219000. | 1902. | 440.3 | 221.2 | 149.0 | 115.5 | 96.71 | 84.88 | 76.82 | 70.99 | 63.19 | 56.36 | 50.38 | 47.08 | 45.14 | 42.76 | 40.54 | 38.45 | 36.49 |
| 4 | 173400. | 1384. | 304.5 | 147.9 | 97.32 | 74.13 | 61.25 | 53.19 | 47.73 | 43.80 | 38.55 | 33.99 | 30.02 | 28.24 | 26.57 | 25.01 | 23.55 | 22.18 | 20.91 |
| 5 | 152400. | 1153. | 245.5 | 116.6 | 75.45 | 56.75 | 46.43 | 40.00 | 35.66 | 32.54 | 28.40 | 24.81 | 21.70 | 20.31 | 19.01 | 17.79 | 16.66 | 15.61 | 14.63 |
| 6 | 140300. | 1025. | 213.3 | 99.63 | 63.67 | 47.44 | 38.51 | 32.98 | 29.24 | 26.57 | 23.03 | 19.97 | 17.33 | 16.14 | 15.04 | 14.01 | 13.06 | 12.17 | 11.34 |
| 7 | 132600. | 944.4 | 193.1 | 89.09 | 56.40 | 41.70 | 33.65 | 28.67 | 25.31 | 22.92 | 19.75 | 17.02- | 14.66 | 13.61 | 12.63 | 11.72 | 10.87 | 10.09 | 9.356 |
| 8 | 127200. | 889.0 | 179.4 | 81.94 | 51.48 | 37.83 | 30.38 | 25.77 | 22.68 | 20.47 | 17.55 | 15.04 | 12.88 | 11.92 | 11.02 | 10.19 | 9.420 | 8.702 | 8.035 |
| 9 | 123300. | 848.7 | 169.5 | 76.80 | 47.95 | 35.06 | 28.03 | 23.70 | 20.79 | 18.72 | 15.99 | 13.64 | 11.62 | 10.72 | 9.881 | 9.105 | 8.385 | 7.716 | 7.094 |
| 10 | 120300. | 818.1 | 162.0 | 72.92 | 45.29 | 32.98 | 26.28 | 22.15 | 19.38 | 17.41 | 14.81 | 12.59 | 10.67 | 9.818 | 9.027 | 8.293 | 7.612 | 6.990 | 6.392 |
| 12 | 116000. | 774.8 | 151.4 | 67.48 | 41.57 | 30.06 | 23.82 | 19.98 | 17.41 | 15.59 | 13.18 | 11.12 | 9.355 | 8.569 | 7.840 | 7.164 | 6.537 | 5.985 | 5.415 |
| 15 | 111900. | 734.4 | 141.6 | 62.45 | 38.14 | 27.39 | 21.57 | 18.00 | 15.61 | 13.92 | 11.69 | 9.782 | 8.151 | 7.425 | 6.753 | 6.130 | 5.553 | 5.016 | 4.517 |
| 20 | 108100. | 696.6 | 132.5 | 57.80 | 34.98 | 24.92 | 19.49 | 16.17 | 13.95 | 12.38 | 10.32 | 8.555 | 7.048 | 6.377 | 5.756 | 5.181 | 4.646 | 4.149 | 3.686 |
| 24 | 106200. | 678.7 | 128.2 | 55.61 | 33.49 | 23.76 | 18.52 | 15.31 | 13.18 | 11.66 | 9.676 | 7.980 | 6.530 | 5.885 | 5.288 | 4.734 | 4.219 | 3.739 | 3.290 |
| 30 | 104500. | 661.4 | 124.1 | 53.50 | 32.07 | 22.65 | 17.59 | 14.49 | 12.43 | 10.98 | 9.061 | 7.428 | 6.033 | 5.413 | 4.838 | 4.303 | 3.805 | 3.340 | 2.903 |
| 40 | 102700. | 644.6 | 120.1 | 51.47 | 30.69 | 21.59 | 16.69 | 13.71 | 11.72 | 10.32 | 8.471 | 6.899 | 5.556 | 4.958 | 4.403 | 3.886 | 3.404 | 2.950 | 2.519 |
| 60 | 101000. | 628.4 | 116.3 | 49.52 | 29.37 | 20.56 | 15.83 | 12.95 | 11.04 | 9.682 | 7.905 | 6.391 | 5.097 | 4.520 | 3.983 | 3.482 | 3.011 | 2.563 | 2.129 |
| 120 | 99410. | 612.8 | 112.5 | 47.64 | 28.11 | 19.58 | 15.01 | 12.23 | 10.38 | 9.075 | 7.362 | 5.903 | 4.654 | 4.095 | 3.574 | 3.085 | 2.621 | 2.170 | 1.710 |
| ∞ | 97820. | 597.7 | 109.0 | 45.84 | 26.89 | 18.63 | 14.22 | 11.53 | 9.752 | 8.492 | 6.840 | 5.433 | 4.225 | 3.683 | 3.175 | 2.693 | 2.225 | 1.748 | 1.000 |

### Explanation

Let (a) $s_1^2$ be an estimate of $\sigma_1^2$ such that $\frac{n_1 s_1^2}{\sigma_1^2}$ is distributed as $\chi^2$ for $n_1$ degrees of freedom

(b) $s_2^2$ be an estimate of $\sigma_2^2$ such that $\frac{n_2 s_2^2}{\sigma^2}$ is distributed independently as $\chi^2$ for $n_2$ degrees of freedom

(c) $F_\epsilon(n_1,n_2)$ be the value of the variance ratio for $n_1$ and $n_2$ degrees of freedom exceeded with probability $\epsilon$

(d) the decision $\frac{\sigma_1^2}{\sigma_2^2} > 1$, or the decision $\frac{\sigma_1^2}{\sigma_2^2} \le 1$, be made according as $\frac{s_1^2}{s_2^2}$ exceeds or does not exceed $F_{0.05}(n_1,n_2)$

Then the probability is $\beta$ that the decision $\frac{\sigma_1^2}{\sigma_2^2} \le 1$ will be made when $\frac{\sigma_1^2}{\sigma_2^2} = \phi(0.05,\beta,n_1,n_2) = \frac{F_{0.05}(n_1,n_2)}{F_{1-\beta}(n_1,n_2)} = F_{0.05}(n_1,n_2) \cdot P_\beta(n_2,n_1)$.

TABLE 8.3    COMPARING TWO STANDARD DEVIATIONS    285

$\beta = 0.01$

| $n_1$ / $n_2$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 15 | 20 | 24 | 30 | 40 | 60 | 120 | ∞ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 654200. | 19650. | 7359. | 4761. | 3742. | 3216. | 2899. | 2690. | 2540. | 2429. | 2276. | 2136. | 2008. | 1948. | 1891. | 1837. | 1785. | 1735. | 1687. |
| 2 | 92560. | 1881. | 590.6 | 346.4 | 256.1 | 211.2 | 184.8 | 167.5 | 155.5 | 146.6 | 134.5 | 122.5 | 113.7 | 106.2 | 104.9 | 100.8 | 96.95 | 93.27 | 89.78 |
| 3 | 54720. | 947.2 | 273.3 | 152.2 | 108.7 | 87.43 | 75.11 | 67.14 | 61.61 | 57.57 | 52.05 | 47.14 | 42.77 | 40.76 | 38.86 | 37.06 | 35.37 | 33.76 | 32.24 |
| 4 | 43360. | 689.2 | 189.2 | 102.1 | 71.27 | 56.38 | 47.82 | 42.32 | 38.52 | 35.75 | 31.99 | 28.96 | 25.71 | 24.36 | 23.09 | 21.89 | 20.76 | 19.69 | 18.68 |
| 5 | 38900. | 574.6 | 152.7 | 80.59 | 55.39 | 43.29 | 36.38 | 31.95 | 28.91 | 26.69 | 23.99 | 21.04 | 18.70 | 17.63 | 16.63 | 15.68 | 14.80 | 13.96 | 13.17 |
| 6 | 35090. | 510.9 | 132.8 | 68.94 | 46.82 | 36.27 | 30.25 | 26.42 | 23.78 | 21.87 | 19.25 | 17.01 | 15.00 | 14.09 | 13.23 | 12.42 | 11.66 | 10.95 | 10.28 |
| 7 | 33150. | 470.7 | 120.3 | 61.71 | 41.53 | 31.93 | 26.48 | 23.02 | 20.64 | 18.91 | 16.58 | 14.54 | 12.74 | 11.92 | 11.16 | 10.43 | 9.758 | 9.122 | 8.524 |
| 8 | 31810. | 443.1 | 111.8 | 56.80 | 37.94 | 29.01 | 23.94 | 20.73 | 18.52 | 16.93 | 14.78 | 12.89 | 11.23 | 10.48 | 9.770 | 9.107 | 8.485 | 7.901 | 7.352 |
| 9 | 30620. | 423.0 | 105.6 | 53.28 | 35.37 | 26.91 | 22.12 | 19.09 | 17.01 | 15.51 | 13.48 | 11.71 | 10.15 | 9.444 | 8.782 | 8.160 | 7.577 | 7.030 | 6.516 |
| 10 | 30060. | 407.8 | 101.0 | 50.59 | 33.43 | 25.33 | 20.76 | 17.86 | 15.88 | 14.44 | 12.51 | 10.62 | 9.843 | 8.672 | 8.042 | 7.452 | 6.898 | 6.378 | 5.890 |
| 12 | 28990. | 386.3 | 94.42 | 46.85 | 30.71 | 23.12 | 18.85 | 16.14 | 14.29 | 12.96 | 11.16 | 9.564 | 8.219 | 7.596 | 7.012 | 6.465 | 5.951 | 5.499 | 5.017 |
| 15 | 27970. | 366.1 | 88.34 | 43.38 | 28.21 | 21.09 | 17.09 | 14.56 | 12.84 | 11.59 | 9.925 | 8.466 | 7.187 | 6.609 | 6.057 | 5.558 | 5.081 | 4.633 | 4.211 |
| 20 | 27020. | 347.4 | 82.70 | 40.18 | 25.90 | 19.22 | 15.47 | 13.11 | 11.50 | 10.34 | 8.798 | 7.439 | 6.240 | 5.702 | 5.197 | 4.723 | 4.278 | 3.858 | 3.462 |
| 24 | 26560. | 338.4 | 80.03 | 38.67 | 24.81 | 18.34 | 14.72 | 12.43 | 10.88 | 9.756 | 8.254 | 6.943 | 5.795 | 5.275 | 4.787 | 4.329 | 3.897 | 3.490 | 3.104 |
| 30 | 26110. | 329.8 | 77.46 | 37.22 | 23.76 | 17.50 | 13.99 | 11.78 | 10.28 | 9.193 | 7.742 | 6.476 | 5.367 | 4.864 | 4.392 | 3.948 | 3.529 | 3.131 | 2.752 |
| 40 | 25680. | 321.5 | 74.97 | 35.82 | 22.76 | 16.69 | 13.29 | 11.15 | 9.700 | 8.652 | 7.251 | 6.027 | 4.955 | 4.468 | 4.011 | 3.579 | 3.170 | 2.779 | 2.403 |
| 60 | 25260. | 313.4 | 72.58 | 34.47 | 21.79 | 15.91 | 12.62 | 10.55 | 9.146 | 8.134 | 6.779 | 5.596 | 4.558 | 4.096 | 3.641 | 3.220 | 2.817 | 2.429 | 2.046 |
| 120 | 24850. | 305.6 | 70.28 | 33.18 | 20.87 | 15.16 | 11.97 | 9.973 | 8.614 | 7.635 | 6.325 | 5.181 | 4.175 | 3.715 | 3.281 | 2.867 | 2.467 | 2.072 | 1.661 |
| ∞ | 24450. | 298.1 | 68.05 | 31.93 | 19.97 | 14.44 | 11.35 | 9.418 | 8.103 | 7.156 | 5.889 | 4.780 | 3.802 | 3.354 | 2.927 | 2.516 | 2.110 | 1.686 | 1.000 |

For explanation see the panel of this table for $\beta = 0.005$.

TABLE 8.3.—$\phi(0.05,\beta,n_1,n_2)$.—(Continued)

$\beta = 0.025$

| $n_2$\\$n_1$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 15 | 20 | 24 | 30 | 40 | 60 | 120 | ∞ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 104600. | 7682. | 3763. | 2744. | 2303. | 2062. | 1911. | 1800. | 1734. | 1678. | 1599. | 1525. | 1456. | 1424. | 1392. | 1362. | 1333. | 1305. | 1278. |
| 2 | 14800. | 741.0 | 307.5 | 205.0 | 162.7 | 140.3 | 126.6 | 117.4 | 110.8 | 105.8 | 98.93 | 92.58 | 86.75 | 84.02 | 81.39 | 78.88 | 76.46 | 74.14 | 71.92 |
| 3 | 8752. | 374.1 | 143.2 | 90.98 | 69.98 | 59.00 | 52.34 | 47.91 | 44.75 | 42.40 | 39.13 | 36.14 | 33.42 | 32.14 | 30.93 | 29.76 | 28.65 | 27.50 | 26.57 |
| 4 | 6935. | 272.5 | 99.54 | 61.36 | 46.22 | 38.38 | 33.66 | 30.52 | 28.30 | 26.65 | 24.36 | 22.28 | 20.39 | 19.51 | 18.67 | 17.87 | 17.11 | 16.38 | 15.68 |
| 5 | 6091. | 227.4 | 80.52 | 48.62 | 36.09 | 29.64 | 25.77 | 23.21 | 21.40 | 20.06 | 18.20 | 16.52 | 14.99 | 14.28 | 13.61 | 12.96 | 12.35 | 11.76 | 11.20 |
| 6 | 5611. | 202.3 | 70.10 | 41.70 | 30.61 | 24.93 | 21.53 | 19.29 | 17.71 | 16.53 | 14.91 | 13.45 | 12.12 | 11.50 | 10.92 | 10.36 | 9.826 | 9.319 | 8.835 |
| 7 | 5302. | 186.4 | 63.57 | 37.39 | 27.22 | 22.02 | 18.92 | 16.87 | 15.43 | 14.36 | 12.89 | 11.56 | 10.36 | 9.801 | 9.270 | 8.765 | 8.283 | 7.825 | 7.388 |
| 8 | 5087. | 175.6 | 59.12 | 34.46 | 24.92 | 20.05 | 17.15 | 15.24 | 13.90 | 12.90 | 11.53 | 10.29 | 9.176 | 8.657 | 8.164 | 7.695 | 7.248 | 6.822 | 6.417 |
| 9 | 4929. | 167.7 | 55.90 | 32.35 | 23.26 | 18.63 | 15.88 | 14.07 | 12.80 | 11.86 | 10.56 | 9.387 | 8.329 | 7.839 | 7.373 | 6.929 | 6.506 | 6.104 | 5.721 |
| 10 | 4809. | 161.6 | 53.47 | 30.76 | 22.01 | 17.57 | 14.93 | 13.19 | 11.97 | 11.07 | 9.827 | 8.706 | 7.694 | 7.225 | 6.779 | 6.355 | 5.950 | 5.565 | 5.198 |
| 12 | 4637. | 153.1 | 50.04 | 28.52 | 20.26 | 16.08 | 13.59 | 11.96 | 10.82 | 9.970 | 8.805 | 7.755 | 6.806 | 6.367 | 5.949 | 5.551 | 5.172 | 4.810 | 4.465 |
| 15 | 4474. | 145.2 | 46.86 | 26.45 | 18.65 | 14.70 | 12.36 | 10.83 | 9.754 | 8.958 | 7.865 | 6.879 | 5.989 | 5.576 | 5.194 | 4.810 | 4.453 | 4.112 | 3.786 |
| 20 | 4321. | 137.8 | 43.90 | 24.53 | 17.16 | 13.43 | 11.23 | 9.787 | 8.774 | 8.027 | 6.999 | 6.072 | 5.235 | 4.847 | 4.476 | 4.112 | 3.785 | 3.461 | 3.149 |
| 24 | 4248. | 134.3 | 42.50 | 23.63 | 16.45 | 12.83 | 10.70 | 9.296 | 8.313 | 7.588 | 6.591 | 5.692 | 4.879 | 4.502 | 4.142 | 3.797 | 3.467 | 3.149 | 2.843 |
| 30 | 4177. | 130.9 | 41.15 | 22.76 | 15.78 | 12.26 | 10.18 | 8.825 | 7.871 | 7.167 | 6.200 | 5.327 | 4.537 | 4.169 | 3.818 | 3.481 | 3.158 | 2.845 | 2.541 |
| 40 | 4108. | 127.6 | 39.85 | 21.92 | 15.13 | 11.71 | 9.691 | 8.372 | 7.446 | 6.762 | 5.823 | 4.975 | 4.206 | 3.848 | 3.504 | 3.174 | 2.855 | 2.545 | 2.238 |
| 60 | 4040. | 124.4 | 38.59 | 21.11 | 14.50 | 11.18 | 9.217 | 7.936 | 7.037 | 6.373 | 5.460 | 4.635 | 3.887 | 3.536 | 3.199 | 2.855 | 2.545 | 2.245 | 1.929 |
| 120 | 3975. | 121.3 | 37.38 | 20.33 | 13.90 | 10.67 | 8.762 | 7.517 | 6.644 | 5.999 | 5.111 | 4.308 | 3.576 | 3.233 | 2.901 | 2.574 | 2.259 | 1.937 | 1.590 |
| ∞ | 3912. | 118.3 | 36.21 | 19.59 | 13.32 | 10.18 | 8.324 | 7.114 | 6.266 | 5.638 | 4.775 | 3.992 | 3.275 | 2.936 | 2.607 | 2.282 | 1.954 | 1.601 | 1.000 |

For explanation see the panel of this table for $\beta = 0.005$.

TABLE 8.3    COMPARING TWO STANDARD DEVIATIONS    287

$$\beta = 0.05$$

| $n_2$ \ $n_1$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 15 | 20 | 24 | 30 | 40 | 60 | 120 | ∞ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 26070. | 3603. | 2185. | 1731. | 1521. | 1401. | 1324. | 1270. | 1231. | 1201. | 1158. | 1117. | 1079. | 1061. | 1043. | 1026. | 1009. | 992.8 | 977.0 |
| 2 | 3603. | 361.0 | 183.1 | 133.7 | 111.6 | 99.42 | 91.68 | 86.38 | 82.51 | 79.58 | 75.43 | 71.54 | 67.92 | 66.20 | 64.53 | 62.92 | 61.37 | 59.86 | 58.40 |
| 3 | 2185. | 183.1 | 86.06 | 60.10 | 48.76 | 42.53 | 38.63 | 35.97 | 34.04 | 32.58 | 30.52 | 28.61 | 26.83 | 25.99 | 25.18 | 24.40 | 23.64 | 22.91 | 22.21 |
| 4 | 1731. | 133.7 | 60.10 | 40.81 | 32.48 | 27.94 | 25.11 | 23.18 | 21.79 | 20.74 | 19.27 | 17.90 | 16.63 | 16.03 | 15.45 | 14.90 | 14.36 | 13.85 | 13.35 |
| 5 | 1521. | 111.6 | 48.76 | 32.48 | 25.51 | 21.72 | 19.36 | 17.77 | 16.62 | 15.75 | 14.53 | 13.40 | 12.36 | 11.86 | 11.39 | 10.93 | 10.49 | 10.07 | 9.665 |
| 6 | 1401. | 99.42 | 42.53 | 27.94 | 21.72 | 18.35 | 16.26 | 14.85 | 13.83 | 13.06 | 11.98 | 10.99 | 10.07 | 9.635 | 9.218 | 8.816 | 8.430 | 8.058 | 7.699 |
| 7 | 1324. | 91.68 | 38.63 | 25.11 | 19.36 | 16.26 | 14.34 | 13.04 | 12.11 | 11.40 | 10.41 | 9.502 | 8.560 | 8.262 | 7.890 | 7.513 | 7.159 | 6.818 | 6.491 |
| 8 | 1270. | 86.38 | 35.97 | 23.18 | 17.77 | 14.85 | 13.04 | 11.82 | 10.94 | 10.28 | 9.355 | 8.499 | 7.709 | 7.337 | 6.979 | 6.634 | 6.302 | 5.982 | 5.675 |
| 9 | 1231. | 82.51 | 34.04 | 21.79 | 16.62 | 13.83 | 12.11 | 10.94 | 10.11 | 9.476 | 8.593 | 7.779 | 7.086 | 6.672 | 6.331 | 6.002 | 5.686 | 5.382 | 5.088 |
| 10 | 1201. | 79.58 | 32.58 | 20.74 | 15.75 | 13.06 | 11.40 | 10.28 | 9.476 | 8.870 | 8.021 | 7.237 | 6.543 | 6.172 | 5.844 | 5.527 | 5.223 | 4.929 | 4.646 |
| 12 | 1158. | 75.43 | 30.52 | 19.27 | 14.53 | 11.98 | 10.41 | 9.355 | 8.593 | 8.021 | 7.218 | 6.476 | 5.746 | 5.471 | 5.160 | 4.860 | 4.571 | 4.293 | 4.023 |
| 15 | 1117. | 71.54 | 28.61 | 17.90 | 13.40 | 10.99 | 9.502 | 8.499 | 7.779 | 7.237 | 6.478 | 5.777 | 5.128 | 4.832 | 4.527 | 4.242 | 3.967 | 3.701 | 3.442 |
| 20 | 1079. | 67.92 | 26.83 | 16.63 | 12.36 | 10.07 | 8.659 | 7.709 | 7.026 | 6.513 | 5.793 | 5.128 | 4.512 | 4.221 | 3.929 | 3.656 | 3.402 | 3.145 | 2.895 |
| 24 | 1061. | 66.20 | 25.99 | 16.03 | 11.86 | 9.636 | 8.262 | 7.337 | 6.672 | 6.172 | 5.471 | 4.832 | 4.221 | 3.935 | 3.660 | 3.392 | 3.132 | 2.879 | 2.630 |
| 30 | 1043. | 64.53 | 25.18 | 15.45 | 11.39 | 9.218 | 7.890 | 6.979 | 6.331 | 5.844 | 5.160 | 4.527 | 3.929 | 3.660 | 3.389 | 3.126 | 2.869 | 2.617 | 2.367 |
| 40 | 1026. | 62.92 | 24.40 | 14.90 | 10.93 | 8.816 | 7.513 | 6.634 | 6.002 | 5.527 | 4.860 | 4.242 | 3.656 | 3.392 | 3.126 | 2.866 | 2.610 | 2.357 | 2.103 |
| 60 | 1009. | 61.37 | 23.64 | 14.36 | 10.49 | 8.430 | 7.159 | 6.302 | 5.686 | 5.223 | 4.571 | 3.967 | 3.402 | 3.132 | 2.869 | 2.610 | 2.354 | 2.097 | 1.831 |
| 120 | 992.8 | 59.86 | 22.91 | 13.85 | 10.07 | 8.058 | 6.818 | 5.982 | 5.382 | 4.929 | 4.293 | 3.701 | 3.145 | 2.879 | 2.617 | 2.357 | 2.097 | 1.838 | 1.532 |
| ∞ | 977.0 | 58.40 | 22.21 | 13.35 | 9.665 | 7.699 | 6.491 | 5.675 | 5.088 | 4.646 | 4.023 | 3.442 | 2.895 | 2.630 | 2.367 | 2.103 | 1.831 | 1.532 | 1.000 |

For explanation see the panel of this table for $\beta = 0.005$.

TABLE 8.3.—$\phi(0.05,\beta,n_1,n_2)$.—(Continued)

$\beta = .10$

| $n_2$\$n_1$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 15 | 20 | 24 | 30 | 40 | 60 | 120 | ∞ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 6436. | 1701. | 1195. | 1021. | 934.5 | 883.5 | 849.9 | 826.0 | 808.3 | 794.6 | 774.8 | 755.9 | 737.8 | 729.0 | 720.4 | 712.1 | 704.0 | 695.9 | 688.1 |
| 2 | 916.4 | 171.0 | 104.7 | 83.24 | 72.93 | 66.95 | 63.04 | 60.30 | 58.28 | 56.72 | 54.49 | 52.37 | 50.35 | 49.38 | 48.44 | 47.52 | 46.62 | 45.74 | 44.89 |
| 3 | 542.8 | 87.51 | 50.01 | 38.21 | 32.62 | 29.40 | 27.32 | 25.86 | 24.79 | 23.96 | 22.78 | 21.67 | 20.61 | 20.11 | 19.61 | 19.13 | 18.66 | 18.21 | 17.77 |
| 4 | 430.4 | 64.19 | 35.22 | 26.24 | 22.02 | 19.60 | 18.04 | 16.95 | 16.15 | 15.54 | 14.66 | 13.83 | 13.05 | 12.67 | 12.31 | 11.95 | 11.61 | 11.27 | 10.95 |
| 5 | 378.2 | 53.77 | 28.72 | 21.03 | 17.44 | 15.38 | 14.06 | 13.14 | 12.46 | 11.94 | 11.20 | 10.50 | 9.837 | 9.521 | 9.213 | 8.913 | 8.622 | 8.339 | 8.063 |
| 6 | 348.5 | 47.96 | 25.14 | 18.18 | 14.94 | 13.09 | 11.89 | 11.06 | 10.46 | 9.990 | 9.324 | 8.696 | 8.102 | 7.818 | 7.541 | 7.273 | 7.011 | 6.757 | 6.509 |
| 7 | 330.4 | 44.29 | 22.89 | 16.39 | 13.38 | 11.65 | 10.55 | 9.777 | 9.211 | 8.779 | 8.160 | 7.577 | 7.026 | 6.762 | 6.505 | 6.255 | 6.012 | 5.775 | 5.545 |
| 8 | 316.1 | 41.77 | 21.35 | 15.18 | 12.31 | 10.68 | 9.632 | 8.902 | 8.367 | 7.957 | 7.371 | 6.818 | 6.296 | 6.046 | 5.802 | 5.565 | 5.334 | 5.109 | 4.890 |
| 9 | 306.3 | 39.93 | 20.24 | 14.30 | 11.55 | 9.979 | 8.972 | 8.272 | 7.757 | 7.364 | 6.802 | 6.271 | 5.770 | 5.529 | 5.295 | 5.067 | 4.844 | 4.628 | 4.416 |
| 10 | 298.8 | 38.53 | 19.40 | 13.63 | 10.97 | 9.449 | 8.474 | 7.796 | 7.298 | 6.917 | 6.373 | 5.859 | 5.372 | 5.139 | 4.912 | 4.690 | 4.474 | 4.263 | 4.057 |
| 12 | 288.2 | 36.55 | 18.20 | 12.70 | 10.15 | 8.703 | 7.773 | 7.127 | 6.652 | 6.289 | 5.769 | 5.279 | 4.814 | 4.590 | 4.372 | 4.159 | 3.952 | 3.748 | 3.549 |
| 15 | 278.1 | 34.70 | 17.10 | 11.83 | 9.394 | 8.012 | 7.124 | 6.507 | 6.054 | 5.707 | 5.210 | 4.740 | 4.294 | 4.079 | 3.870 | 3.664 | 3.464 | 3.266 | 3.072 |
| 20 | 268.6 | 32.98 | 16.06 | 11.02 | 8.693 | 7.372 | 6.523 | 5.933 | 5.499 | 5.167 | 4.691 | 4.240 | 3.810 | 3.603 | 3.400 | 3.200 | 3.004 | 2.811 | 2.618 |
| 24 | 264.1 | 32.16 | 15.57 | 10.64 | 8.361 | 7.069 | 6.239 | 5.662 | 5.237 | 4.912 | 4.445 | 4.003 | 3.581 | 3.376 | 3.176 | 2.978 | 2.783 | 2.590 | 2.397 |
| 30 | 259.7 | 31.36 | 15.10 | 10.27 | 8.042 | 6.777 | 5.965 | 5.400 | 4.984 | 4.666 | 4.208 | 3.773 | 3.358 | 3.156 | 2.957 | 2.761 | 2.567 | 2.373 | 2.177 |
| 40 | 255.4 | 30.59 | 14.65 | 9.912 | 7.734 | 6.497 | 5.701 | 5.148 | 4.741 | 4.428 | 3.979 | 3.551 | 3.141 | 2.942 | 2.744 | 2.549 | 2.353 | 2.156 | 1.954 |
| 60 | 251.3 | 29.85 | 14.21 | 9.569 | 7.437 | 6.226 | 5.447 | 4.905 | 4.506 | 4.199 | 3.758 | 3.336 | 2.931 | 2.733 | 2.536 | 2.339 | 2.141 | 1.937 | 1.723 |
| 120 | 247.2 | 29.13 | 13.78 | 9.239 | 7.151 | 5.965 | 5.202 | 4.670 | 4.279 | 3.977 | 3.543 | 3.128 | 2.726 | 2.528 | 2.330 | 2.130 | 1.926 | 1.710 | 1.465 |
| ∞ | 243.3 | 28.43 | 13.37 | 8.920 | 6.875 | 5.713 | 4.965 | 4.444 | 4.059 | 3.763 | 3.335 | 2.925 | 2.524 | 2.326 | 2.125 | 1.919 | 1.702 | 1.457 | 1.000 |

For explanation see the panel of this table for $\beta = 0.005$.

TABLE 8.3    COMPARING TWO STANDARD DEVIATIONS    289

$$\beta = 0.25$$

| $\frac{m}{n}$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 15 | 20 | 24 | 30 | 40 | 60 | 120 | ∞ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 941.0 | 513.0 | 436.6 | 405.9 | 389.5 | 379.4 | 372.5 | 367.5 | 363.7 | 360.8 | 356.4 | 352.3 | 348.1 | 346.1 | 344.1 | 342.2 | 340.3 | 338.4 | 336.5 |
| 2 | 138.8 | 57.00 | 43.69 | 38.49 | 35.75 | 34.06 | 32.92 | 32.10 | 31.47 | 30.99 | 30.27 | 29.58 | 28.92 | 28.59 | 28.27 | 27.95 | 27.64 | 27.33 | 27.03 |
| 3 | 63.05 | 30.12 | 21.85 | 18.66 | 16.98 | 15.95 | 15.26 | 14.76 | 14.38 | 14.08 | 13.65 | 13.23 | 12.82 | 12.63 | 12.43 | 12.24 | 12.05 | 11.86 | 11.68 |
| 4 | 66.15 | 22.44 | 15.75 | 13.19 | 11.84 | 11.01 | 10.46 | 10.05 | 9.750 | 9.513 | 9.165 | 8.828 | 8.502 | 8.342 | 8.184 | 8.030 | 7.876 | 7.726 | 7.577 |
| 5 | 68.28 | 18.98 | 13.08 | 10.76 | 9.569 | 8.837 | 8.343 | 7.986 | 7.717 | 7.507 | 7.199 | 6.900 | 6.609 | 6.467 | 6.327 | 6.188 | 6.052 | 5.917 | 5.784 |
| 6 | 63.79 | 17.04 | 11.52 | 9.415 | 8.312 | 7.634 | 7.176 | 6.846 | 6.596 | 6.401 | 6.114 | 5.836 | 5.566 | 5.433 | 5.302 | 5.173 | 5.045 | 4.919 | 4.794 |
| 7 | 50.89 | 15.80 | 10.55 | 8.566 | 7.520 | 6.877 | 6.442 | 6.128 | 5.891 | 5.705 | 5.422 | 5.167 | 4.909 | 4.782 | 4.657 | 4.533 | 4.411 | 4.289 | 4.170 |
| 8 | 48.88 | 14.95 | 9.907 | 7.985 | 6.978 | 6.359 | 5.940 | 5.637 | 5.408 | 5.229 | 4.946 | 4.709 | 4.459 | 4.336 | 4.214 | 4.094 | 3.975 | 3.857 | 3.740 |
| 9 | 47.40 | 14.33 | 9.429 | 7.562 | 6.584 | 5.983 | 5.575 | 5.280 | 5.057 | 4.883 | 4.628 | 4.376 | 4.131 | 4.011 | 3.892 | 3.774 | 3.657 | 3.540 | 3.425 |
| 10 | 46.27 | 13.86 | 9.066 | 7.241 | 6.285 | 5.697 | 5.298 | 5.010 | 4.791 | 4.620 | 4.368 | 4.123 | 3.882 | 3.764 | 3.646 | 3.530 | 3.414 | 3.299 | 3.185 |
| 12 | 44.65 | 13.18 | 8.551 | 6.788 | 5.863 | 5.294 | 4.907 | 4.627 | 4.415 | 4.248 | 4.004 | 3.764 | 3.527 | 3.413 | 3.297 | 3.183 | 3.068 | 2.955 | 2.841 |
| 15 | 43.13 | 12.56 | 8.071 | 6.365 | 5.469 | 4.917 | 4.542 | 4.270 | 4.064 | 3.902 | 3.662 | 3.428 | 3.197 | 3.083 | 2.969 | 2.855 | 2.741 | 2.627 | 2.513 |
| 20 | 41.60 | 11.97 | 7.623 | 5.970 | 5.102 | 4.566 | 4.201 | 3.937 | 3.735 | 3.577 | 3.343 | 3.113 | 2.885 | 2.771 | 2.658 | 2.544 | 2.429 | 2.313 | 2.196 |
| 24 | 41.00 | 11.69 | 7.409 | 5.782 | 4.927 | 4.399 | 4.040 | 3.778 | 3.579 | 3.422 | 3.191 | 2.962 | 2.735 | 2.621 | 2.508 | 2.393 | 2.277 | 2.160 | 2.039 |
| 30 | 40.33 | 11.42 | 7.203 | 5.601 | 4.759 | 4.238 | 3.883 | 3.625 | 3.428 | 3.273 | 3.043 | 2.815 | 2.589 | 2.475 | 2.361 | 2.245 | 2.127 | 2.007 | 1.883 |
| 40 | 39.68 | 11.15 | 7.004 | 5.426 | 4.596 | 4.082 | 3.732 | 3.476 | 3.282 | 3.127 | 2.899 | 2.673 | 2.446 | 2.331 | 2.216 | 2.099 | 1.978 | 1.853 | 1.721 |
| 60 | 39.05 | 10.90 | 6.812 | 5.257 | 4.439 | 3.932 | 3.585 | 3.333 | 3.140 | 2.987 | 2.760 | 2.533 | 2.306 | 2.191 | 2.073 | 1.953 | 1.828 | 1.695 | 1.551 |
| 120 | 38.43 | 10.65 | 6.625 | 5.093 | 4.287 | 3.786 | 3.443 | 3.193 | 3.002 | 2.850 | 2.624 | 2.398 | 2.169 | 2.051 | 1.931 | 1.806 | 1.674 | 1.530 | 1.359 |
| ∞ | 37.84 | 10.41 | 6.445 | 4.935 | 4.139 | 3.645 | 3.306 | 3.058 | 2.868 | 2.717 | 2.492 | 2.265 | 2.033 | 1.913 | 1.788 | 1.656 | 1.512 | 1.342 | 1.000 |

For explanation see the panel of this table for $\beta = 0.005$.

TABLE 8.3.—$\phi(0.05, \beta, n_1, n_2)$.—(Continued)

$\beta = 0.50$

| $n_2$ / $n_1$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 15 | 20 | 24 | 30 | 40 | 60 | 120 | $\infty$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 161.5 | 133.0 | 126.2 | 123.2 | 121.5 | 120.5 | 119.7 | 119.2 | 118.8 | 118.5 | 118.0 | 117.5 | 117.0 | 116.8 | 116.6 | 116.4 | 116.1 | 115.9 | 115.7 |
| 2 | 27.77 | 19.00 | 16.89 | 15.94 | 15.41 | 15.07 | 14.84 | 14.66 | 14.53 | 14.42 | 14.26 | 14.11 | 13.96 | 13.88 | 13.81 | 13.73 | 13.66 | 13.59 | 13.51 |
| 3 | 17.31 | 10.84 | 9.277 | 8.576 | 8.176 | 7.920 | 7.740 | 7.607 | 7.506 | 7.425 | 7.304 | 7.186 | 7.068 | 7.011 | 6.963 | 6.895 | 6.838 | 6.781 | 6.724 |
| 4 | 14.06 | 8.382 | 7.008 | 6.388 | 6.035 | 5.905 | 5.644 | 5.525 | 5.434 | 5.361 | 5.253 | 5.145 | 5.038 | 4.985 | 4.933 | 4.890 | 4.828 | 4.775 | 4.723 |
| 5 | 12.51 | 7.244 | 5.963 | 5.383 | 5.050 | 4.834 | 4.682 | 4.569 | 4.482 | 4.413 | 4.309 | 4.207 | 4.104 | 4.053 | 4.002 | 3.951 | 3.901 | 3.849 | 3.799 |
| 6 | 11.63 | 6.596 | 5.371 | 4.813 | 4.493 | 4.284 | 4.137 | 4.027 | 3.942 | 3.875 | 3.773 | 3.673 | 3.572 | 3.522 | 3.472 | 3.422 | 3.371 | 3.321 | 3.270 |
| 7 | 11.06 | 6.180 | 4.991 | 4.449 | 4.136 | 3.931 | 3.787 | 3.679 | 3.596 | 3.529 | 3.430 | 3.330 | 3.230 | 3.180 | 3.130 | 3.080 | 3.029 | 2.979 | 2.928 |
| 8 | 10.66 | 5.892 | 4.728 | 4.196 | 3.888 | 3.687 | 3.545 | 3.438 | 3.356 | 3.290 | 3.191 | 3.091 | 2.991 | 2.941 | 2.891 | 2.841 | 2.790 | 2.739 | 2.688 |
| 9 | 10.36 | 5.680 | 4.535 | 4.011 | 3.707 | 3.508 | 3.367 | 3.261 | 3.179 | 3.113 | 3.014 | 2.915 | 2.816 | 2.765 | 2.715 | 2.664 | 2.613 | 2.561 | 2.509 |
| 10 | 10.14 | 5.518 | 4.388 | 3.870 | 3.569 | 3.371 | 3.231 | 3.126 | 3.044 | 2.978 | 2.880 | 2.780 | 2.680 | 2.630 | 2.579 | 2.528 | 2.476 | 2.424 | 2.371 |
| 12 | 9.815 | 5.288 | 4.178 | 3.668 | 3.371 | 3.176 | 3.037 | 2.932 | 2.851 | 2.785 | 2.687 | 2.587 | 2.486 | 2.435 | 2.384 | 2.331 | 2.278 | 2.225 | 2.170 |
| 15 | 9.509 | 5.071 | 3.981 | 3.479 | 3.186 | 2.992 | 2.854 | 2.750 | 2.668 | 2.603 | 2.504 | 2.404 | 2.301 | 2.249 | 2.197 | 2.143 | 2.088 | 2.032 | 1.975 |
| 20 | 9.220 | 4.866 | 3.796 | 3.301 | 3.011 | 2.819 | 2.681 | 2.577 | 2.496 | 2.430 | 2.330 | 2.228 | 2.124 | 2.071 | 2.016 | 1.960 | 1.903 | 1.844 | 1.783 |
| 24 | 9.082 | 4.769 | 3.708 | 3.216 | 2.927 | 2.736 | 2.598 | 2.494 | 2.413 | 2.347 | 2.246 | 2.144 | 2.038 | 1.984 | 1.928 | 1.871 | 1.812 | 1.750 | 1.686 |
| 30 | 8.947 | 4.674 | 3.622 | 3.133 | 2.846 | 2.655 | 2.518 | 2.414 | 2.332 | 2.266 | 2.165 | 2.061 | 1.954 | 1.898 | 1.841 | 1.782 | 1.720 | 1.656 | 1.596 |
| 40 | 8.817 | 4.582 | 3.538 | 3.053 | 2.767 | 2.577 | 2.439 | 2.335 | 2.253 | 2.186 | 2.085 | 1.980 | 1.870 | 1.813 | 1.754 | 1.693 | 1.628 | 1.559 | 1.494 |
| 60 | 8.688 | 4.493 | 3.458 | 2.975 | 2.691 | 2.500 | 2.363 | 2.259 | 2.176 | 2.109 | 2.006 | 1.900 | 1.788 | 1.729 | 1.668 | 1.603 | 1.534 | 1.459 | 1.374 |
| 120 | 8.564 | 4.406 | 3.379 | 2.900 | 2.617 | 2.426 | 2.289 | 2.184 | 2.101 | 2.034 | 1.930 | 1.821 | 1.706 | 1.645 | 1.581 | 1.512 | 1.437 | 1.352 | 1.247 |
| $\infty$ | 8.444 | 4.322 | 3.303 | 2.826 | 2.544 | 2.354 | 2.217 | 2.112 | 2.028 | 1.960 | 1.854 | 1.743 | 1.624 | 1.560 | 1.492 | 1.418 | 1.333 | 1.228 | 1.000 |

For explanation see the panel of this table for $\beta = 0.005$.

TABLE 8.3   COMPARING TWO STANDARD DEVIATIONS   291

$\beta = 0.75$

| $n_2 \backslash n_1$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 15 | 20 | 24 | 30 | 40 | 60 | 120 | ∞ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 27.70 | 26.60 | 26.31 | 26.17 | 26.10 | 26.05 | 26.01 | 25.99 | 25.97 | 25.95 | 25.93 | 25.91 | 25.88 | 25.87 | 25.86 | 25.85 | 25.84 | 25.83 | 25.82 |
| 2 | 7.200 | 6.333 | 6.077 | 5.955 | 5.883 | 5.836 | 5.803 | 5.778 | 5.759 | 5.744 | 5.721 | 5.698 | 5.676 | 5.664 | 5.653 | 5.642 | 5.631 | 5.620 | 5.609 |
| 3 | 5.004 | 4.190 | 3.938 | 3.815 | 3.741 | 3.692 | 3.657 | 3.630 | 3.610 | 3.594 | 3.560 | 3.545 | 3.520 | 3.508 | 3.496 | 3.483 | 3.471 | 3.458 | 3.446 |
| 4 | 4.265 | 3.472 | 3.221 | 3.095 | 3.019 | 2.968 | 2.931 | 2.904 | 2.882 | 2.865 | 2.831 | 2.812 | 2.786 | 2.773 | 2.759 | 2.746 | 2.732 | 2.719 | 2.705 |
| 5 | 3.904 | 3.123 | 2.871 | 2.743 | 2.665 | 2.613 | 2.576 | 2.546 | 2.524 | 2.505 | 2.478 | 2.450 | 2.422 | 2.408 | 2.393 | 2.379 | 2.364 | 2.350 | 2.335 |
| 6 | 3.693 | 2.919 | 2.666 | 2.537 | 2.458 | 2.404 | 2.365 | 2.335 | 2.312 | 2.293 | 2.264 | 2.236 | 2.205 | 2.190 | 2.175 | 2.160 | 2.144 | 2.128 | 2.112 |
| 7 | 3.554 | 2.786 | 2.532 | 2.402 | 2.321 | 2.266 | 2.226 | 2.196 | 2.172 | 2.152 | 2.125 | 2.094 | 2.061 | 2.045 | 2.029 | 2.013 | 1.997 | 1.980 | 1.963 |
| 8 | 3.457 | 2.691 | 2.437 | 2.306 | 2.225 | 2.169 | 2.128 | 2.097 | 2.072 | 2.052 | 2.024 | 1.990 | 1.963 | 1.942 | 1.925 | 1.908 | 1.891 | 1.874 | 1.856 |
| 9 | 3.384 | 2.622 | 2.368 | 2.235 | 2.153 | 2.097 | 2.056 | 2.023 | 1.998 | 1.978 | 1.946 | 1.914 | 1.881 | 1.864 | 1.847 | 1.829 | 1.811 | 1.793 | 1.774 |
| 10 | 3.329 | 2.568 | 2.314 | 2.181 | 2.098 | 2.041 | 1.999 | 1.966 | 1.941 | 1.920 | 1.888 | 1.846 | 1.821 | 1.803 | 1.786 | 1.767 | 1.749 | 1.729 | 1.710 |
| 12 | 3.249 | 2.491 | 2.236 | 2.102 | 2.018 | 1.960 | 1.917 | 1.884 | 1.858 | 1.836 | 1.803 | 1.770 | 1.733 | 1.715 | 1.696 | 1.676 | 1.656 | 1.636 | 1.615 |
| 15 | 3.172 | 2.418 | 2.163 | 2.027 | 1.942 | 1.883 | 1.839 | 1.805 | 1.778 | 1.755 | 1.721 | 1.696 | 1.649 | 1.628 | 1.608 | 1.587 | 1.566 | 1.543 | 1.520 |
| 20 | 3.100 | 2.349 | 2.092 | 1.956 | 1.870 | 1.809 | 1.764 | 1.729 | 1.701 | 1.678 | 1.642 | 1.604 | 1.564 | 1.543 | 1.522 | 1.499 | 1.475 | 1.450 | 1.424 |
| 24 | 3.055 | 2.316 | 2.059 | 1.922 | 1.835 | 1.773 | 1.728 | 1.692 | 1.663 | 1.640 | 1.603 | 1.564 | 1.525 | 1.501 | 1.479 | 1.455 | 1.430 | 1.403 | 1.375 |
| 30 | 3.031 | 2.283 | 2.028 | 1.888 | 1.800 | 1.738 | 1.692 | 1.656 | 1.627 | 1.603 | 1.565 | 1.525 | 1.482 | 1.459 | 1.436 | 1.411 | 1.384 | 1.355 | 1.324 |
| 40 | 2.998 | 2.251 | 1.994 | 1.855 | 1.767 | 1.704 | 1.657 | 1.620 | 1.591 | 1.566 | 1.527 | 1.486 | 1.441 | 1.417 | 1.392 | 1.365 | 1.337 | 1.306 | 1.270 |
| 60 | 2.945 | 2.220 | 1.962 | 1.824 | 1.734 | 1.671 | 1.623 | 1.586 | 1.555 | 1.530 | 1.490 | 1.447 | 1.401 | 1.375 | 1.349 | 1.320 | 1.288 | 1.252 | 1.211 |
| 120 | 2.934 | 2.190 | 1.932 | 1.792 | 1.702 | 1.638 | 1.590 | 1.551 | 1.520 | 1.494 | 1.453 | 1.409 | 1.360 | 1.333 | 1.304 | 1.272 | 1.237 | 1.196 | 1.141 |
| ∞ | 2.903 | 2.161 | 1.902 | 1.762 | 1.671 | 1.606 | 1.557 | 1.517 | 1.486 | 1.459 | 1.416 | 1.370 | 1.318 | 1.289 | 1.258 | 1.222 | 1.181 | 1.127 | 1.000 |

For explanation see the panel of this table for $\beta = 0.005$.

TABLE 8.3—$\phi(0.05,\beta,n_1,n_2)$.—(Continued)

$\beta = 0.90$

| $n_2$ \ $n_1$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 15 | 20 | 24 | 30 | 40 | 60 | 120 | ∞ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 4.050 | 4.030 | 4.025 | 4.022 | 4.021 | 4.020 | 4.019 | 4.019 | 4.019 | 4.018 | 4.018 | 4.017 | 4.017 | 4.017 | 4.017 | 4.016 | 4.016 | 4.016 | 4.016 |
| 2 | 2.171 | 2.111 | 2.092 | 2.083 | 2.076 | 2.073 | 2.070 | 2.068 | 2.067 | 2.065 | 2.063 | 2.061 | 2.060 | 2.059 | 2.058 | 2.057 | 2.056 | 2.055 | 2.054 |
| 3 | 1.829 | 1.749 | 1.721 | 1.706 | 1.698 | 1.692 | 1.688 | 1.684 | 1.682 | 1.680 | 1.677 | 1.674 | 1.670 | 1.669 | 1.667 | 1.666 | 1.664 | 1.662 | 1.661 |
| 4 | 1.696 | 1.606 | 1.573 | 1.555 | 1.544 | 1.537 | 1.532 | 1.527 | 1.524 | 1.522 | 1.518 | 1.514 | 1.509 | 1.507 | 1.505 | 1.503 | 1.501 | 1.499 | 1.497 |
| 5 | 1.627 | 1.531 | 1.495 | 1.475 | 1.463 | 1.454 | 1.448 | 1.443 | 1.439 | 1.436 | 1.431 | 1.426 | 1.421 | 1.419 | 1.416 | 1.414 | 1.411 | 1.408 | 1.406 |
| 6 | 1.586 | 1.485 | 1.446 | 1.425 | 1.412 | 1.402 | 1.395 | 1.390 | 1.386 | 1.382 | 1.377 | 1.372 | 1.366 | 1.363 | 1.360 | 1.357 | 1.354 | 1.351 | 1.348 |
| 7 | 1.558 | 1.454 | 1.414 | 1.392 | 1.377 | 1.367 | 1.360 | 1.354 | 1.349 | 1.346 | 1.340 | 1.334 | 1.328 | 1.324 | 1.321 | 1.318 | 1.314 | 1.311 | 1.307 |
| 8 | 1.538 | 1.432 | 1.391 | 1.368 | 1.352 | 1.342 | 1.334 | 1.328 | 1.323 | 1.319 | 1.313 | 1.306 | 1.299 | 1.296 | 1.292 | 1.289 | 1.285 | 1.281 | 1.277 |
| 9 | 1.523 | 1.416 | 1.373 | 1.349 | 1.334 | 1.323 | 1.314 | 1.308 | 1.303 | 1.298 | 1.292 | 1.285 | 1.278 | 1.274 | 1.270 | 1.266 | 1.262 | 1.258 | 1.254 |
| 10 | 1.511 | 1.403 | 1.359 | 1.335 | 1.319 | 1.307 | 1.299 | 1.292 | 1.287 | 1.282 | 1.275 | 1.268 | 1.261 | 1.257 | 1.253 | 1.248 | 1.244 | 1.239 | 1.235 |
| 12 | 1.494 | 1.384 | 1.340 | 1.314 | 1.297 | 1.285 | 1.276 | 1.269 | 1.263 | 1.259 | 1.251 | 1.243 | 1.235 | 1.231 | 1.226 | 1.221 | 1.217 | 1.212 | 1.206 |
| 15 | 1.478 | 1.366 | 1.320 | 1.294 | 1.276 | 1.264 | 1.254 | 1.247 | 1.240 | 1.235 | 1.227 | 1.219 | 1.210 | 1.205 | 1.200 | 1.194 | 1.189 | 1.183 | 1.177 |
| 20 | 1.463 | 1.349 | 1.302 | 1.274 | 1.256 | 1.243 | 1.233 | 1.224 | 1.218 | 1.212 | 1.204 | 1.194 | 1.184 | 1.179 | 1.173 | 1.167 | 1.161 | 1.154 | 1.147 |
| 24 | 1.455 | 1.341 | 1.293 | 1.265 | 1.246 | 1.232 | 1.222 | 1.214 | 1.207 | 1.201 | 1.192 | 1.182 | 1.171 | 1.166 | 1.160 | 1.153 | 1.146 | 1.139 | 1.131 |
| 30 | 1.448 | 1.332 | 1.284 | 1.256 | 1.236 | 1.222 | 1.211 | 1.203 | 1.196 | 1.190 | 1.180 | 1.170 | 1.159 | 1.152 | 1.146 | 1.139 | 1.131 | 1.123 | 1.114 |
| 40 | 1.441 | 1.324 | 1.275 | 1.246 | 1.227 | 1.212 | 1.201 | 1.192 | 1.185 | 1.178 | 1.168 | 1.158 | 1.146 | 1.139 | 1.132 | 1.124 | 1.116 | 1.107 | 1.096 |
| 60 | 1.433 | 1.316 | 1.267 | 1.237 | 1.217 | 1.202 | 1.191 | 1.182 | 1.174 | 1.167 | 1.157 | 1.145 | 1.132 | 1.125 | 1.118 | 1.109 | 1.100 | 1.089 | 1.076 |
| 120 | 1.427 | 1.309 | 1.258 | 1.228 | 1.208 | 1.193 | 1.181 | 1.171 | 1.163 | 1.156 | 1.145 | 1.133 | 1.119 | 1.111 | 1.103 | 1.093 | 1.082 | 1.069 | 1.051 |
| ∞ | 1.420 | 1.301 | 1.250 | 1.220 | 1.199 | 1.183 | 1.171 | 1.161 | 1.152 | 1.145 | 1.134 | 1.121 | 1.106 | 1.097 | 1.087 | 1.076 | 1.063 | 1.045 | 1.000 |

For explanation see the panel of this table for $\beta = 0.005$.

TABLE 8.3    COMPARING TWO STANDARD DEVIATIONS    293

$\beta = 0.975$

| $n_2 \backslash n_1$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 15 | 20 | 24 | 30 | 40 | 60 | 120 | ∞ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.2492 | 0.2495 | 0.2496 | 0.2496 | 0.2497 | 0.2497 | 0.2497 | 0.2497 | 0.2497 | 0.2497 | 0.2497 | 0.2497 | 0.2497 | 0.2497 | 0.2497 | 0.2497 | 0.2498 | 0.2498 | 0.2498 |
| 2 | 0.4808 | 0.4872 | 0.4893 | 0.4904 | 0.4910 | 0.4915 | 0.4918 | 0.4920 | 0.4922 | 0.4923 | 0.4925 | 0.4927 | 0.4930 | 0.4931 | 0.4931 | 0.4933 | 0.4934 | 0.4935 | 0.4936 |
| 3 | 0.5906 | 0.5954 | 0.6009 | 0.6037 | 0.6055 | 0.6068 | 0.6077 | 0.6083 | 0.6089 | 0.6093 | 0.6099 | 0.6106 | 0.6113 | 0.6116 | 0.6119 | 0.6123 | 0.6126 | 0.6130 | 0.6133 |
| 4 | 0.6309 | 0.6521 | 0.6605 | 0.6651 | 0.6681 | 0.6701 | 0.6716 | 0.6727 | 0.6737 | 0.6744 | 0.6755 | 0.6767 | 0.6779 | 0.6785 | 0.6791 | 0.6797 | 0.6803 | 0.6809 | 0.6816 |
| 5 | 0.6603 | 0.6861 | 0.6968 | 0.7028 | 0.7067 | 0.7094 | 0.7115 | 0.7131 | 0.7143 | 0.7154 | 0.7169 | 0.7186 | 0.7202 | 0.7211 | 0.7220 | 0.7229 | 0.7238 | 0.7247 | 0.7256 |
| 6 | 0.6794 | 0.7065 | 0.7209 | 0.7280 | 0.7327 | 0.7361 | 0.7386 | 0.7406 | 0.7421 | 0.7434 | 0.7454 | 0.7475 | 0.7496 | 0.7507 | 0.7518 | 0.7530 | 0.7542 | 0.7554 | 0.7566 |
| 7 | 0.6926 | 0.7242 | 0.7380 | 0.7461 | 0.7514 | 0.7553 | 0.7582 | 0.7604 | 0.7623 | 0.7638 | 0.7661 | 0.7686 | 0.7712 | 0.7725 | 0.7738 | 0.7752 | 0.7767 | 0.7782 | 0.7797 |
| 8 | 0.7024 | 0.7359 | 0.7508 | 0.7596 | 0.7655 | 0.7697 | 0.7730 | 0.7755 | 0.7776 | 0.7793 | 0.7820 | 0.7847 | 0.7877 | 0.7892 | 0.7908 | 0.7924 | 0.7941 | 0.7959 | 0.7977 |
| 9 | 0.7098 | 0.7448 | 0.7606 | 0.7700 | 0.7764 | 0.7810 | 0.7845 | 0.7873 | 0.7896 | 0.7915 | 0.7944 | 0.7975 | 0.8008 | 0.8025 | 0.8043 | 0.8061 | 0.8080 | 0.8100 | 0.8121 |
| 10 | 0.7157 | 0.7519 | 0.7685 | 0.7784 | 0.7851 | 0.7901 | 0.7938 | 0.7968 | 0.7993 | 0.8013 | 0.8045 | 0.8078 | 0.8114 | 0.8133 | 0.8153 | 0.8174 | 0.8195 | 0.8217 | 0.8240 |
| 12 | 0.7243 | 0.7624 | 0.7801 | 0.7908 | 0.7982 | 0.8036 | 0.8078 | 0.8112 | 0.8139 | 0.8162 | 0.8198 | 0.8236 | 0.8278 | 0.8300 | 0.8323 | 0.8347 | 0.8372 | 0.8399 | 0.8427 |
| 15 | 0.7328 | 0.7728 | 0.7916 | 0.8032 | 0.8112 | 0.8172 | 0.8218 | 0.8256 | 0.8286 | 0.8312 | 0.8353 | 0.8398 | 0.8446 | 0.8471 | 0.8499 | 0.8527 | 0.8558 | 0.8590 | 0.8624 |
| 20 | 0.7411 | 0.7829 | 0.8030 | 0.8155 | 0.8242 | 0.8308 | 0.8359 | 0.8401 | 0.8436 | 0.8465 | 0.8512 | 0.8568 | 0.8610 | 0.8650 | 0.8682 | 0.8717 | 0.8754 | 0.8795 | 0.8839 |
| 24 | 0.7451 | 0.7879 | 0.8086 | 0.8215 | 0.8307 | 0.8376 | 0.8430 | 0.8474 | 0.8511 | 0.8542 | 0.8592 | 0.8647 | 0.8708 | 0.8742 | 0.8778 | 0.8816 | 0.8858 | 0.8904 | 0.8955 |
| 30 | 0.7492 | 0.7929 | 0.8141 | 0.8276 | 0.8371 | 0.8444 | 0.8501 | 0.8548 | 0.8587 | 0.8620 | 0.8674 | 0.8733 | 0.8800 | 0.8837 | 0.8877 | 0.8919 | 0.8967 | 0.9020 | 0.9080 |
| 40 | 0.7531 | 0.7978 | 0.8197 | 0.8336 | 0.8436 | 0.8512 | 0.8572 | 0.8621 | 0.8663 | 0.8698 | 0.8756 | 0.8820 | 0.8933 | 0.8934 | 0.8978 | 0.9027 | 0.9082 | 0.9144 | 0.9217 |
| 60 | 0.7570 | 0.8026 | 0.8252 | 0.8396 | 0.8500 | 0.8579 | 0.8642 | 0.8695 | 0.8739 | 0.8777 | 0.8839 | 0.8909 | 0.8989 | 0.9035 | 0.9085 | 0.9142 | 0.9205 | 0.9281 | 0.9373 |
| 120 | 0.7608 | 0.8074 | 0.8306 | 0.8455 | 0.8564 | 0.8647 | 0.8713 | 0.8769 | 0.8817 | 0.8857 | 0.8924 | 0.9000 | 0.9089 | 0.9140 | 0.9198 | 0.9263 | 0.9340 | 0.9436 | 0.9569 |
| ∞ | 0.7646 | 0.8121 | 0.8359 | 0.8514 | 0.8627 | 0.8714 | 0.8785 | 0.8844 | 0.8894 | 0.8938 | 0.9010 | 0.9093 | 0.9192 | 0.9251 | 0.9317 | 0.9397 | 0.9494 | 0.9629 | 1.0000 |

For explanation see the panel of this table for $\beta = 0.005$.

TABLE 8.3—$\phi(0.05,\beta,n_1 n_2)$.—(Continued)

$\beta = 0.99$

| $n_1$ \ $n_2$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 15 | 20 | 24 | 30 | 40 | 60 | 120 | ∞ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.03994 | 0.03990 | 0.03992 | 0.03993 | 0.03993 | 0.03994 | 0.03994 | 0.03994 | 0.03994 | 0.03994 | 0.03994 | 0.03994 | 0.03995 | 0.03995 | 0.03995 | 0.03995 | 0.03995 | 0.03995 | 0.03995 |
| 2 | 0.1879 | 0.1919 | 0.1933 | 0.1939 | 0.1943 | 0.1946 | 0.1948 | 0.1949 | 0.1950 | 0.1951 | 0.1953 | 0.1954 | 0.1955 | 0.1956 | 0.1957 | 0.1957 | 0.1958 | 0.1959 | 0.1959 |
| 3 | 0.2969 | 0.3100 | 0.3149 | 0.3176 | 0.3192 | 0.3203 | 0.3211 | 0.3218 | 0.3223 | 0.3227 | 0.3233 | 0.3239 | 0.3245 | 0.3248 | 0.3251 | 0.3254 | 0.3257 | 0.3261 | 0.3264 |
| 4 | 0.3636 | 0.3858 | 0.3948 | 0.3998 | 0.4030 | 0.4053 | 0.4069 | 0.4082 | 0.4092 | 0.4100 | 0.4113 | 0.4126 | 0.4139 | 0.4146 | 0.4152 | 0.4159 | 0.4166 | 0.4173 | 0.4180 |
| 5 | 0.4064 | 0.4359 | 0.4485 | 0.4558 | 0.4605 | 0.4639 | 0.4663 | 0.4683 | 0.4698 | 0.4711 | 0.4731 | 0.4751 | 0.4772 | 0.4782 | 0.4793 | 0.4804 | 0.4816 | 0.4827 | 0.4839 |
| 6 | 0.4356 | 0.4708 | 0.4864 | 0.4956 | 0.5017 | 0.5060 | 0.5093 | 0.5118 | 0.5139 | 0.5156 | 0.5182 | 0.5210 | 0.5238 | 0.5253 | 0.5268 | 0.5284 | 0.5300 | 0.5316 | 0.5332 |
| 7 | 0.4566 | 0.4962 | 0.5143 | 0.5251 | 0.5323 | 0.5376 | 0.5416 | 0.5447 | 0.5472 | 0.5493 | 0.5526 | 0.5560 | 0.5596 | 0.5615 | 0.5634 | 0.5654 | 0.5674 | 0.5695 | 0.5717 |
| 8 | 0.4723 | 0.5155 | 0.5357 | 0.5478 | 0.5560 | 0.5620 | 0.5666 | 0.5703 | 0.5732 | 0.5757 | 0.5795 | 0.5836 | 0.5878 | 0.5901 | 0.5924 | 0.5948 | 0.5973 | 0.5999 | 0.6025 |
| 9 | 0.4846 | 0.5306 | 0.5524 | 0.5657 | 0.5748 | 0.5815 | 0.5866 | 0.5907 | 0.5941 | 0.5968 | 0.6012 | 0.6058 | 0.6108 | 0.6133 | 0.6160 | 0.6188 | 0.6217 | 0.6247 | 0.6279 |
| 10 | 0.4946 | 0.5427 | 0.5660 | 0.5802 | 0.5901 | 0.5973 | 0.6030 | 0.6075 | 0.6111 | 0.6142 | 0.6190 | 0.6241 | 0.6297 | 0.6326 | 0.6357 | 0.6388 | 0.6421 | 0.6456 | 0.6492 |
| 12 | 0.5088 | 0.5609 | 0.5863 | 0.6022 | 0.6133 | 0.6215 | 0.6280 | 0.6331 | 0.6374 | 0.6409 | 0.6465 | 0.6527 | 0.6592 | 0.6627 | 0.6664 | 0.6703 | 0.6744 | 0.6787 | 0.6832 |
| 15 | 0.5232 | 0.5791 | 0.6069 | 0.6245 | 0.6369 | 0.6462 | 0.6535 | 0.6595 | 0.6644 | 0.6685 | 0.6752 | 0.6824 | 0.6903 | 0.6945 | 0.6990 | 0.7038 | 0.7089 | 0.7143 | 0.7202 |
| 20 | 0.5375 | 0.5982 | 0.6274 | 0.6469 | 0.6606 | 0.6713 | 0.6797 | 0.6865 | 0.6922 | 0.6971 | 0.7049 | 0.7135 | 0.7231 | 0.7283 | 0.7339 | 0.7399 | 0.7464 | 0.7535 | 0.7613 |
| 24 | 0.5445 | 0.6081 | 0.6377 | 0.6581 | 0.6728 | 0.6840 | 0.6930 | 0.7003 | 0.7064 | 0.7117 | 0.7202 | 0.7296 | 0.7402 | 0.7460 | 0.7523 | 0.7591 | 0.7665 | 0.7748 | 0.7840 |
| 30 | 0.5515 | 0.6151 | 0.6480 | 0.6694 | 0.6849 | 0.6968 | 0.7064 | 0.7143 | 0.7209 | 0.7266 | 0.7359 | 0.7462 | 0.7579 | 0.7645 | 0.7715 | 0.7793 | 0.7879 | 0.7976 | 0.8086 |
| 40 | 0.5585 | 0.6241 | 0.6582 | 0.6807 | 0.6971 | 0.7098 | 0.7200 | 0.7284 | 0.7356 | 0.7417 | 0.7518 | 0.7632 | 0.7763 | 0.7836 | 0.7917 | 0.8007 | 0.8108 | 0.8223 | 0.8361 |
| 60 | 0.5654 | 0.6329 | 0.6685 | 0.6920 | 0.7093 | 0.7227 | 0.7337 | 0.7427 | 0.7505 | 0.7571 | 0.7682 | 0.7807 | 0.7953 | 0.8037 | 0.8130 | 0.8235 | 0.8355 | 0.8500 | 0.8680 |
| 120 | 0.5722 | 0.6418 | 0.6787 | 0.7033 | 0.7216 | 0.7358 | 0.7474 | 0.7572 | 0.7656 | 0.7728 | 0.7849 | 0.7988 | 0.8152 | 0.8248 | 0.8356 | 0.8482 | 0.8631 | 0.8819 | 0.9063 |
| ∞ | 0.5790 | 0.6505 | 0.6888 | 0.7146 | 0.7338 | 0.7490 | 0.7614 | 0.7719 | 0.7809 | 0.7888 | 0.8020 | 0.8175 | 0.8361 | 0.8473 | 0.8601 | 0.8755 | 0.8948 | 0.9221 | 1.000 |

For explanation see the panel of this table for $\beta = 0.005$.

TABLE 8.3   COMPARING TWO STANDARD DEVIATIONS   295

$$\beta = 0.995$$

| $n_2$ \ $n_1$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 15 | 20 | 24 | 30 | 40 | 60 | 120 | ∞ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.009959 | 0.009975 | 0.009980 | 0.009981 | 0.009983 | 0.009984 | 0.009984 | 0.009985 | 0.009985 | 0.009985 | 0.009986 | 0.009986 | 0.009986 | 0.009986 | 0.009986 | 0.009986 | 0.009987 | 0.009987 | 0.009987 |
| 2 | 0.09326 | 0.09548 | 0.09622 | 0.09650 | 0.09682 | 0.09697 | 0.09708 | 0.09716 | 0.09722 | 0.09727 | 0.09735 | 0.09742 | 0.09750 | 0.09753 | 0.09757 | 0.09761 | 0.09765 | 0.09768 | 0.09772 |
| 3 | 0.1823 | 0.1918 | 0.1954 | 0.1974 | 0.1986 | 0.1994 | 0.2000 | 0.2005 | 0.2008 | 0.2011 | 0.2015 | 0.2020 | 0.2024 | 0.2027 | 0.2029 | 0.2031 | 0.2034 | 0.2036 | 0.2038 |
| 4 | 0.2460 | 0.2642 | 0.2717 | 0.2759 | 0.2786 | 0.2805 | 0.2819 | 0.2829 | 0.2838 | 0.2845 | 0.2854 | 0.2866 | 0.2877 | 0.2883 | 0.2889 | 0.2894 | 0.2900 | 0.2906 | 0.2912 |
| 5 | 0.2900 | 0.3159 | 0.3273 | 0.3338 | 0.3380 | 0.3411 | 0.3434 | 0.3451 | 0.3465 | 0.3477 | 0.3495 | 0.3513 | 0.3535 | 0.3542 | 0.3552 | 0.3562 | 0.3573 | 0.3584 | 0.3594 |
| 6 | 0.3213 | 0.3536 | 0.3683 | 0.3769 | 0.3827 | 0.3869 | 0.3900 | 0.3925 | 0.3945 | 0.3961 | 0.3986 | 0.4013 | 0.4040 | 0.4055 | 0.4069 | 0.4084 | 0.4100 | 0.4116 | 0.4132 |
| 7 | 0.3444 | 0.3819 | 0.3994 | 0.4100 | 0.4171 | 0.4223 | 0.4262 | 0.4293 | 0.4319 | 0.4339 | 0.4372 | 0.4406 | 0.4443 | 0.4461 | 0.4480 | 0.4500 | 0.4521 | 0.4542 | 0.4564 |
| 8 | 0.3620 | 0.4038 | 0.4237 | 0.4359 | 0.4443 | 0.4503 | 0.4550 | 0.4587 | 0.4617 | 0.4642 | 0.4681 | 0.4723 | 0.4768 | 0.4790 | 0.4814 | 0.4839 | 0.4865 | 0.4892 | 0.4920 |
| 9 | 0.3759 | 0.4211 | 0.4431 | 0.4567 | 0.4660 | 0.4729 | 0.4782 | 0.4825 | 0.4860 | 0.4889 | 0.4934 | 0.4983 | 0.5035 | 0.5063 | 0.5091 | 0.5121 | 0.5152 | 0.5194 | 0.5218 |
| 10 | 0.3871 | 0.4352 | 0.4589 | 0.4737 | 0.4839 | 0.4916 | 0.4975 | 0.5022 | 0.5061 | 0.5094 | 0.5145 | 0.5200 | 0.5260 | 0.5291 | 0.5324 | 0.5338 | 0.5394 | 0.5432 | 0.5471 |
| 12 | 0.4039 | 0.4566 | 0.4830 | 0.4998 | 0.5116 | 0.5204 | 0.5274 | 0.5329 | 0.5376 | 0.5414 | 0.5476 | 0.5543 | 0.5616 | 0.5654 | 0.5695 | 0.5737 | 0.5703 | 0.5631 | 0.5682 |
| 15 | 0.4207 | 0.4782 | 0.5076 | 0.5266 | 0.5401 | 0.5503 | 0.5584 | 0.5650 | 0.5704 | 0.5750 | 0.5826 | 0.5906 | 0.5995 | 0.6043 | 0.6094 | 0.6149 | 0.6207 | 0.6269 | 0.6336 |
| 20 | 0.4376 | 0.4999 | 0.5326 | 0.5539 | 0.5693 | 0.5812 | 0.5906 | 0.5983 | 0.6048 | 0.6103 | 0.6193 | 0.6292 | 0.6402 | 0.6463 | 0.6528 | 0.6599 | 0.6675 | 0.6759 | 0.6851 |
| 24 | 0.4460 | 0.5109 | 0.5452 | 0.5678 | 0.5842 | 0.5969 | 0.6071 | 0.6155 | 0.6225 | 0.6286 | 0.6384 | 0.6494 | 0.6618 | 0.6687 | 0.6761 | 0.6842 | 0.6930 | 0.7029 | 0.7139 |
| 30 | 0.4544 | 0.5218 | 0.5578 | 0.5817 | 0.5993 | 0.6129 | 0.6239 | 0.6330 | 0.6407 | 0.6473 | 0.6582 | 0.6703 | 0.6843 | 0.6921 | 0.7005 | 0.7099 | 0.7203 | 0.7321 | 0.7455 |
| 40 | 0.4627 | 0.5327 | 0.5705 | 0.5958 | 0.6145 | 0.6291 | 0.6410 | 0.6508 | 0.6592 | 0.6665 | 0.6784 | 0.6920 | 0.7077 | 0.7166 | 0.7264 | 0.7373 | 0.7497 | 0.7640 | 0.7811 |
| 60 | 0.4710 | 0.5436 | 0.5832 | 0.6100 | 0.6299 | 0.6455 | 0.6583 | 0.6690 | 0.6782 | 0.6861 | 0.6993 | 0.7144 | 0.7322 | 0.7425 | 0.7539 | 0.7669 | 0.7819 | 0.8000 | 0.8228 |
| 120 | 0.4793 | 0.5545 | 0.5960 | 0.6242 | 0.6454 | 0.6621 | 0.6759 | 0.6875 | 0.6975 | 0.7062 | 0.7208 | 0.7378 | 0.7581 | 0.7699 | 0.7835 | 0.7992 | 0.8180 | 0.8420 | 0.8762 |
| ∞ | 0.4875 | 0.5654 | 0.6087 | 0.6384 | 0.6609 | 0.6789 | 0.6937 | 0.7063 | 0.7172 | 0.7268 | 0.7430 | 0.7620 | 0.7853 | 0.7993 | 0.8165 | 0.8352 | 0.8600 | 0.8957 | 1.000 |

For explanation see the panel of this table for $\beta = 0.005$.

Since, under the circumstances assumed, the assignment of the designations I and II to the products is arbitrary, it is natural to require that (26) be the reciprocal of (25). For this to be the case, it is necessary that

$$F^2_{0.50}(n_1,n_2) = F_\beta(n_1,n_2) \cdot F_{1-\beta}(n_1,n_2) = \frac{F_\beta(n_1,n_2)}{F_\beta(n_2,n_1)} \qquad (27)$$
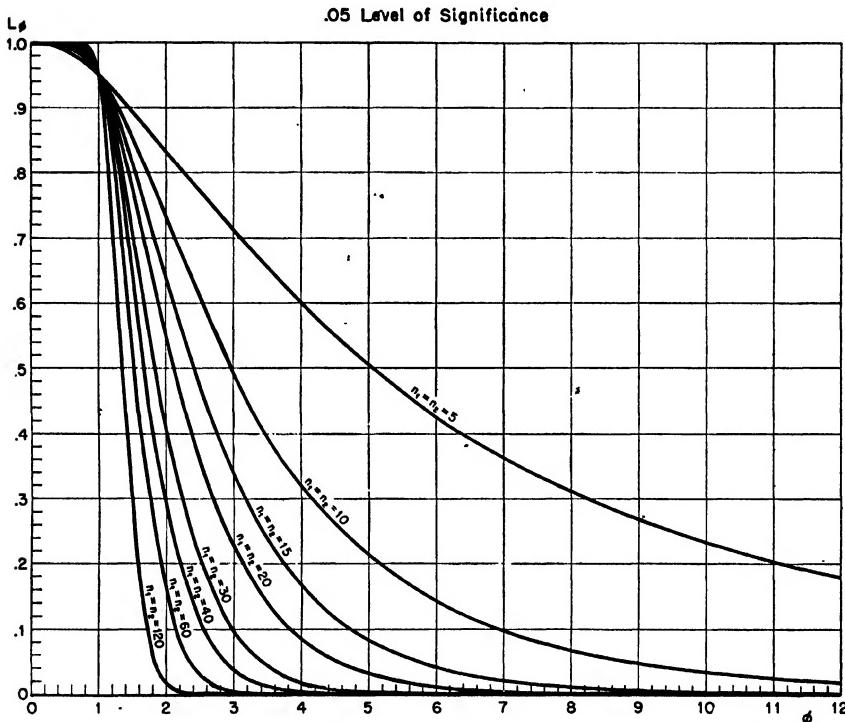


FIG. 8.3.—Operating-characteristic curves for experiments to determine whether a particular one of two new products (processes) is more variable than the other, 0.05 level of significance.

It is clear that when $n_1 = n_2$, equation (27) holds for every value of $\beta$; it is believed that the condition $n_1 = n_2$ is necessary for the equality to hold. Consequently, if symmetry of action is considered desirable in the situation considered in this section, the test should be carried out with the same number of degrees of freedom for $s_1^2$ and $s_2^2$.

Values of

$$\phi(0.50,\beta,n,n) = F_\beta(n,n) = \phi(\beta,0.50,n,n) \qquad (28)$$

can be obtained from Tables 8.3 and 8.4 by means of (27). For example, $\phi(0.50,0.05,30,30) = \phi(0.05,0.50,30,30) = 1.841$, from Table 8.3.

### 5. EXPERIMENTS INVOLVING THE ANALYSIS OF VARIANCE

### 5.1. Systematic and Random Variation

Many experimer ts are analyzed by a technique (due to R. A. Fisher) known as the analysis of variance.*    Tables 8.3 and 8.4 are useful in
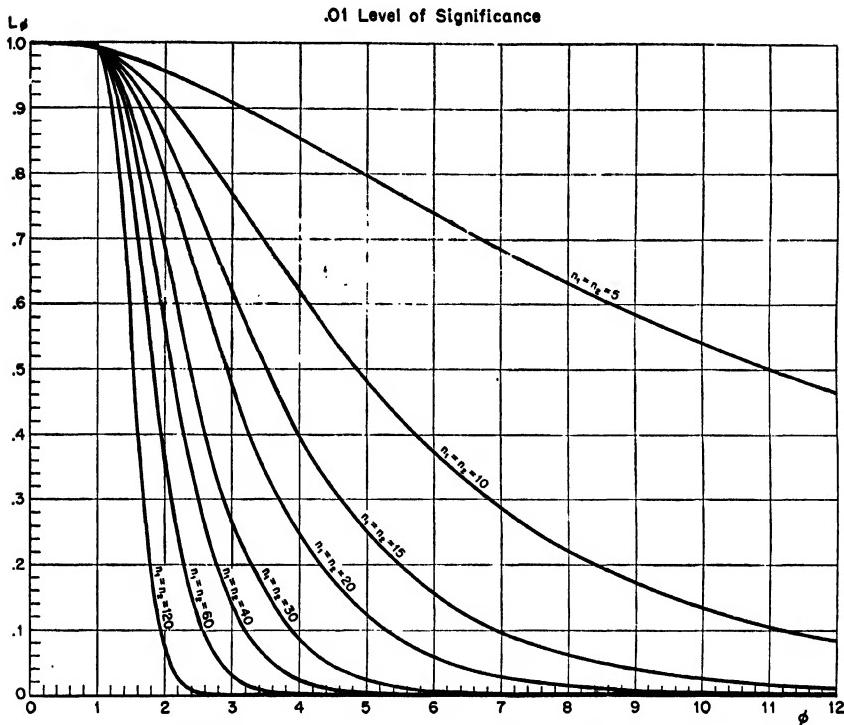


FIG. 8.4.—Operating-characteristic curves for experiments to determine whether a particular one of two new products (processes) is more variable than the other, 0.01 level of significance.

planning a certain class of such experiments, namely, those for detecting and measuring components of random variation.

The factors associated with the variation of a product can be divided into two categories:

(1) There are factors that introduce *random* variation.    Lack of perfect control at some stage of a production process often acts in this

* Excellent accounts of the analysis of variance, which are especially pertinent to this chapter, are given by Baines, *Economical Design*, and Daniels, "Components of Variance."

## TABLE 8.4

### VALUES OF $\phi(0.01,\beta,n_1,n_2)$ CORRESPONDING TO SELECTED VALUES OF $\beta$, $n_1$, AND $n_2$

$$\beta = 0.005$$

| $n_2$ \\ $n_1$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 15 | 20 | 24 | 30 | 40 | 60 | 120 | ∞ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 65690000. | 962400. | 300200. | 176300. | 131300. | 109200. | 96250. | 87860. | 81990. | 77670. | 71770. | 66490. | 61740. | 59550. | 57470. | 55500. | 53630. | 51850. | 50160. |
| 2 | 1970000. | 19700. | 4938. | 2609. | 1819. | 1445. | 1232. | 1097. | 1005. | 937.0 | 846.0 | 765.7 | 694.8 | 662.5 | 632.4 | 603.4 | 576.5 | 551.1 | 527.2 |
| 3 | 737400. | 6138. | 1398. | 696.5 | 466.8 | 360.5 | 301.1 | 263.8 | 238.4 | 220.0 | 195.5 | 174.0 | 155.3 | 146.8 | 138.9 | 131.4 | 124.4 | 117.9 | 111.8 |
| 4 | 477000. | 3566. | 771.2 | 369.9 | 241.5 | 182.9 | 150.5 | 130.3 | 116.6 | 106.8 | 93.73 | 82.39 | 72.54 | 68.11 | 63.98 | 60.12 | 56.52 | 53.16 | 50.02 |
| 5 | 374800. | 2646. | 547.4 | 255.8 | 163.8 | 122.3 | 99.56 | 85.42 | 75.89 | 69.07 | 60.03 | 52.23 | 45.49 | 42.46 | 39.65 | 37.03 | 34.60 | 32.33 | 30.22 |
| 6 | 322100. | 2178. | 438.5 | 201.0 | 126.9 | 93.75 | 75.62 | 64.42 | 56.90 | 51.53 | 44.43 | 38.33 | 33.07 | 30.73 | 28.55 | 26.52 | 24.64 | 22.89 | 21.27 |
| 7 | 290400. | 1903. | 375.5 | 169.7 | 105.9 | 77.57 | 62.13 | 52.63 | 46.26 | 41.72 | 35.74 | 30.61 | 26.20 | 24.24 | 22.42 | 20.73 | 19.17 | 17.71 | 16.37 |
| 8 | 269400. | 1724. | 335.0 | 149.6 | 92.59 | 67.31 | 53.61 | 45.19 | 39.56 | 35.56 | 30.29 | 25.78 | 21.92 | 20.20 | 18.61 | 17.14 | 15.77 | 14.51 | 13.33 |
| 9 | 254400. | 1509. | 306.8 | 135.8 | 83.42 | 60.29 | 47.79 | 40.12 | 35.00 | 31.37 | 26.59 | 22.51 | 19.02 | 17.47 | 16.04 | 14.71 | 13.49 | 12.35 | 11.30 |
| 10 | 243900. | 1507. | 288.2 | 125.7 | 76.76 | 55.30 | 43.58 | 36.46 | 31.72 | 28.35 | 23.93 | 20.16 | 16.95 | 15.52 | 14.20 | 12.98 | 11.85 | 10.81 | 9.846 |
| 12 | 227900. | 1381. | 258.3 | 112.1 | 67.78 | 48.37 | 37.93 | 31.56 | 27.32 | 24.32 | 20.39 | 17.04 | 14.19 | 12.93 | 11.76 | 10.69 | 9.694 | 8.775 | 7.926 |
| 15 | 213900. | 1268. | 233.4 | 100.0 | 59.89 | 42.38 | 33.00 | 27.29 | 23.50 | 20.82 | 17.31 | 14.33 | 11.81 | 10.69 | 9.661 | 8.710 | 7.833 | 7.022 | 6.273 |
| 20 | 201100. | 1167. | 211.2 | 89.35 | 52.94 | 37.12 | 28.68 | 23.55 | 20.16 | 17.76 | 14.64 | 11.99 | 9.747 | 8.757 | 7.844 | 7.002 | 6.225 | 5.507 | 4.842 |
| 24 | 195100. | 1120. | 201.1 | 84.49 | 49.78 | 34.72 | 26.73 | 21.87 | 18.65 | 16.39 | 13.43 | 10.94 | 8.822 | 7.899 | 7.029 | 6.236 | 5.504 | 4.825 | 4.197 |
| 30 | 189400. | 1075. | 191.5 | 79.92 | 46.81 | 32.51 | 24.90 | 20.29 | 17.25 | 15.11 | 12.31 | 9.955 | 7.961 | 7.081 | 6.270 | 5.522 | 4.830 | 4.187 | 3.589 |
| 40 | 183900. | 1033. | 182.5 | 75.62 | 44.03 | 30.41 | 23.19 | 18.82 | 15.94 | 13.91 | 11.27 | 9.040 | 7.158 | 6.327 | 5.562 | 4.854 | 4.198 | 3.587 | 3.012 |
| 60 | 178700. | 992.9 | 173.9 | 71.56 | 41.41 | 28.45 | 21.58 | 17.44 | 14.71 | 12.79 | 10.29 | 8.187 | 6.409 | 5.624 | 4.899 | 4.228 | 3.603 | 3.016 | 2.453 |
| 120 | 173700. | 954.0 | 165.8 | 67.74 | 38.95 | 26.61 | 20.08 | 16.15 | 13.56 | 11.74 | 9.380 | 7.390 | 5.709 | 4.965 | 4.277 | 3.638 | 3.037 | 2.461 | 1.883 |
| ∞ | 169000. | 918.8 | 158.2 | 64.14 | 36.64 | 24.88 | 18.68 | 14.94 | 12.49 | 10.77 | 8.529 | 6.646 | 5.053 | 4.347 | 3.691 | 3.076 | 2.487 | 1.896 | 1.000 |

### Explanation

Let (a) $s_1^2$ be an estimate of $\sigma_1^2$ such that $\frac{n_1 s_1^2}{\sigma_1^2}$ is distributed as $\chi^2$ for $n_1$ degrees of freedom

(b) $s_2^2$ be an estimate of $\sigma_2^2$ such that $\frac{n_2 s_2^2}{\sigma_2^2}$ is distributed independently as $\chi^2$ for $n_2$ degrees of freedom

(c) $F_\epsilon(n_1, n_2)$ be the value of the variance ratio for $n_1$ and $n_2$ degrees of freedom exceeded with probability $\epsilon$

(d) the decision $\frac{\sigma_1^2}{\sigma_2^2} > 1$, or the decision $\frac{\sigma_1^2}{\sigma_2^2} \leq 1$, be made according as $\frac{s_1^2}{s_2^2}$ exceeds or does not exceed $F_{0.01}(n_1,n_2)$

Then the probability is $\beta$ that the decision $\frac{\sigma_1^2}{\sigma_2^2} \leq 1$ will be made when $\frac{\sigma_1^2}{\sigma_2^2} = \phi(0.01,\beta,n_1,n_2) = \frac{F_{0.01}(n_1,n_2)}{F_{1-\beta}(n_1,n_2)} = F_{0.01}(n_1,n_2) \cdot F_\beta(n_2,n_1)$.

TABLE 8.4    COMPARING TWO STANDARD DEVIATIONS                299

$$\beta = 0.01$$

| $n_2 \backslash n_1$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 15 | 20 | 24 | 30 | 40 | 60 | 120 | ∞ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1849000000. | 492500. | 184300. | 119200. | 93710. | 80530. | 72600. | 67350. | 63600. | 60820. | 56970. | 54460. | 50270. | 48770. | 47350. | 45980. | 44680. | 43430. | 42240. |
| 2 | 492500. | 9801. | 3056. | 1788. | 1318. | 1085. | 948.5 | 859.5 | 797.2 | 751.4 | 688.6 | 632.3 | 581.7 | 558.3 | 536.2 | 515.1 | 495.2 | 476.2 | 458.2 |
| 3 | 184300. | 3056. | 867.7 | 479.3 | 340.5 | 273.0 | 233.9 | 208.7 | 191.2 | 178.4 | 161.0 | 145.6 | 131.8 | 125.5 | 119.5 | 113.9 | 108.6 | 103.6 | 98.79 |
| 4 | 119200. | 1788. | 479.3 | 255.3 | 176.8 | 139.1 | 117.5 | 103.7 | 94.14 | 87.19 | 77.79 | 69.47 | 62.12 | 58.76 | 55.60 | 52.62 | 49.82 | 47.18 | 44.69 |
| 5 | 93710. | 1318. | 340.5 | 176.8 | 120.3 | 93.34 | 78.01 | 68.23 | 61.53 | 56.65 | 50.08 | 44.29 | 39.19 | 36.87 | 34.69 | 32.65 | 30.72 | 28.92 | 27.22 |
| 6 | 80530. | 1085. | 273.0 | 139.1 | 93.34 | 71.67 | 59.40 | 51.61 | 46.28 | 42.41 | 37.21 | 32.64 | 28.63 | 26.81 | 25.11 | 23.51 | 22.01 | 20.60 | 19.28 |
| 7 | 72600. | 948.5 | 233.9 | 117.5 | 78.01 | 59.40 | 48.90 | 42.26 | 37.71 | 34.43 | 30.01 | 26.15 | 22.77 | 21.24 | 19.80 | 18.46 | 17.20 | 16.02 | 14.91 |
| 8 | 67350. | 859.5 | 208.7 | 103.7 | 68.23 | 51.61 | 42.26 | 36.35 | 32.31 | 29.40 | 25.50 | 22.09 | 19.10 | 17.75 | 16.49 | 15.31 | 14.21 | 13.17 | 12.20 |
| 9 | 63600. | 797.2 | 191.2 | 94.14 | 61.53 | 46.28 | 37.71 | 32.31 | 28.63 | 25.98 | 22.43 | 19.33 | 16.62 | 15.40 | 14.25 | 13.19 | 12.19 | 11.25 | 10.38 |
| 10 | 60820. | 751.4 | 178.4 | 87.19 | 56.65 | 42.41 | 34.43 | 29.40 | 25.98 | 23.51 | 20.22 | 17.34 | 14.84 | 13.71 | 12.65 | 11.66 | 10.74 | 9.880 | 9.072 |
| 12 | 56970. | 688.6 | 161.0 | 77.79 | 50.08 | 37.21 | 30.01 | 25.50 | 22.43 | 20.22 | 17.27 | 14.70 | 12.47 | 11.46 | 10.52 | 9.644 | 8.825 | 8.059 | 7.343 |
| 15 | 54460. | 632.3 | 145.6 | 69.47 | 44.29 | 32.64 | 26.15 | 22.09 | 19.33 | 17.34 | 14.70 | 12.41 | 10.41 | 9.516 | 8.679 | 7.897 | 7.168 | 6.496 | 5.847 |
| 20 | 50270. | 581.7 | 131.8 | 62.12 | 39.19 | 28.63 | 22.77 | 19.10 | 16.62 | 14.84 | 12.47 | 10.41 | 8.950 | 7.839 | 7.062 | 6.383 | 5.731 | 5.121 | 4.548 |
| 24 | 48770. | 558.3 | 125.5 | 58.76 | 36.87 | 26.81 | 21.24 | 17.75 | 15.40 | 13.71 | 11.46 | 9.516 | 7.839 | 7.071 | 6.363 | 5.703 | 5.064 | 4.504 | 3.959 |
| 30 | 47350. | 536.2 | 119.5 | 55.60 | 34.69 | 25.11 | 19.80 | 16.49 | 14.25 | 12.65 | 10.52 | 8.679 | 7.062 | 6.363 | 5.993 | 5.066 | 4.479 | 3.926 | 3.403 |
| 40 | 45980. | 515.1 | 113.9 | 52.62 | 32.65 | 23.51 | 18.46 | 15.31 | 13.19 | 11.66 | 9.644 | 7.897 | 6.383 | 5.703 | 5.066 | 4.470 | 3.910 | 3.380 | 2.874 |
| 60 | 44680. | 495.2 | 108.6 | 49.82 | 30.72 | 22.01 | 17.20 | 14.21 | 12.19 | 10.74 | 8.825 | 7.168 | 5.731 | 5.064 | 4.479 | 3.910 | 3.372 | 2.858 | 2.358 |
| 120 | 43430. | 476.2 | 103.6 | 47.18 | 28.92 | 20.60 | 16.02 | 13.17 | 11.25 | 9.880 | 8.059 | 6.496 | 5.121 | 4.504 | 3.926 | 3.380 | 2.858 | 2.350 | 1.829 |
| ∞ | 42240. | 458.2 | 98.79 | 44.69 | 27.22 | 19.28 | 14.91 | 12.20 | 10.38 | 9.072 | 7.343 | 5.847 | 4.548 | 3.959 | 3.403 | 2.874 | 2.358 | 1.829 | 1.000 |

For explanation see the panel of this table for $\beta = 0.005$,

TABLE 8.4.—$\phi(0.01,\beta,n_1,n_2)$.—(Continued)

$\beta = 0.025$

| $n_1$ \ $n_2$ | ∞ | 120 | 60 | 40 | 30 | 24 | 20 | 15 | 12 | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 31980. | 32660. | 33370. | 34100. | 34860. | 35640. | 36450. | 38170. | 40020. | 42010. | 43420. | 45290. | 47860. | 51640. | 57680. | 68720. | 94250. | 192500. | 2625000. |
| 2 | 367.0 | 378.5 | 390.5 | 403.0 | 416.0 | 429.5 | 443.7 | 473.8 | 506.6 | 542.4 | 568.0 | 602.2 | 649.9 | 721.1 | 837.4 | 1057. | 1691. | 3861. | 78750. |
| 3 | 81.41 | 84.62 | 87.96 | 91.47 | 95.14 | 98.97 | 103.0 | 111.6 | 121.0 | 131.4 | 138.9 | 148.9 | 163.0 | 184.2 | 219.2 | 286.5 | 454.8 | 1207. | 29480. |
| 4 | 37.51 | 39.24 | 41.06 | 42.97 | 44.97 | 47.07 | 49.28 | 54.01 | 59.24 | 65.00 | 69.18 | 74.77 | 82.71 | 94.70 | 114.7 | 153.5 | 252.1 | 706.5 | 19070. |
| 5 | 23.15 | 24.36 | 25.64 | 26.98 | 28.39 | 29.86 | 31.42 | 34.77 | 38.48 | 42.58 | 45.55 | 49.57 | 55.26 | 63.90 | 78.37 | 106.7 | 179.5 | 521.6 | 14990. |
| 6 | 16.57 | 17.53 | 18.54 | 19.60 | 20.72 | 21.90 | 23.14 | 25.81 | 28.78 | 32.06 | 34.45 | 37.69 | 42.28 | 49.27 | 61.03 | 84.14 | 144.1 | 429.7 | 12890. |
| 7 | 12.92 | 13.74 | 14.60 | 15.50 | 16.45 | 17.46 | 18.51 | 20.80 | 23.33 | 26.15 | 28.20 | 30.98 | 34.93 | 40.96 | 51.13 | 71.20 | 123.6 | 375.7 | 11610. |
| 8 | 10.65 | 11.37 | 12.13 | 12.94 | 13.78 | 14.67 | 15.61 | 17.64 | 19.90 | 22.41 | 24.25 | 26.73 | 30.27 | 35.67 | 44.81 | 62.91 | 110.4 | 340.5 | 10770. |
| 9 | 9.111 | 9.771 | 10.47 | 11.20 | 11.97 | 12.78 | 13.64 | 15.50 | 17.56 | 19.86 | 21.54 | 23.82 | 27.07 | 32.05 | 40.47 | 57.19 | 101.2 | 315.9 | 10170. |
| 10 | 8.007 | 8.620 | 9.267 | 9.948 | 10.66 | 11.42 | 12.22 | 13.95 | 15.88 | 18.02 | 19.59 | 21.72 | 24.76 | 29.41 | 37.31 | 53.01 | 94.48 | 297.8 | 9729. |
| 12 | 6.536 | 7.088 | 7.669 | 8.281 | 8.926 | 9.607 | 10.32 | 11.88 | 13.62 | 15.56 | 16.97 | 18.90 | 21.65 | 25.87 | 33.04 | 47.36 | 85.34 | 273.0 | 9113. |
| 15 | 5.257 | 5.756 | 6.281 | 6.833 | 7.416 | 8.029 | 8.676 | 10.08 | 11.65 | 13.40 | 14.68 | 16.42 | 18.92 | 22.75 | 29.28 | 42.36 | 77.21 | 250.7 | 8853. |
| 20 | 4.137 | 4.593 | 5.071 | 5.572 | 6.099 | 6.655 | 7.240 | 8.510 | 9.929 | 11.51 | 12.68 | 14.26 | 16.52 | 20.01 | 25.96 | 37.93 | 69.96 | 230.7 | 8040. |
| 24 | 3.628 | 4.065 | 4.523 | 5.002 | 5.505 | 6.034 | 6.592 | 7.801 | 9.151 | 10.66 | 11.77 | 13.27 | 15.43 | 18.76 | 24.45 | 35.90 | 66.64 | 221.5 | 7801. |
| 30 | 3.142 | 3.567 | 4.008 | 4.467 | 4.948 | 5.454 | 5.986 | 7.139 | 8.425 | 9.864 | 10.92 | 12.35 | 14.42 | 17.59 | 23.03 | 34.00 | 63.50 | 212.7 | 7573. |
| 40 | 2.677 | 3.095 | 3.522 | 3.965 | 4.426 | 4.910 | 5.418 | 6.518 | 7.745 | 9.117 | 10.12 | 11.49 | 13.46 | 16.50 | 21.70 | 32.20 | 60.54 | 204.4 | 7355. |
| 60 | 2.222 | 2.641 | 3.061 | 3.490 | 3.935 | 4.400 | 4.887 | 5.938 | 7.108 | 8.418 | 9.377 | 10.68 | 12.56 | 15.47 | 20.44 | 30.51 | 57.73 | 196.5 | 7146. |
| 120 | 1.751 | 2.196 | 2.618 | 3.039 | 3.472 | 3.919 | 4.387 | 5.394 | 6.512 | 7.762 | 8.678 | 9.927 | 11.72 | 14.50 | 19.26 | 28.91 | 55.08 | 189.0 | 6947. |
| ∞ | 1.000 | 1.736 | 2.183 | 2.607 | 3.031 | 3.466 | 3.917 | 4.883 | 5.953 | 7.148 | 8.023 | 9.217 | 10.93 | 13.59 | 18.15 | 27.41 | 52.57 | 181.9 | 6756. |

For explanation see the panel of this table for $\beta = 0.005$.

TABLE 8.4    COMPARING TWO STANDARD DEVIATIONS    301

β = 0.05

| $n_1$ \ $n_2$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 15 | 20 | 24 | 30 | 40 | 60 | 120 | ∞ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 654200. | 92560. | 54720. | 43360. | 36090. | 35080. | 33150. | 31810. | 30820. | 30060. | 28990. | 27330. | 27020. | 26560. | 26110. | 25680. | 25260. | 24850. | 24450. |
| 2 | 19050. | 1881. | 947.2 | 689.2 | 574.6 | 510.9 | 470.7 | 443.1 | 423.0 | 407.8 | 386.3 | 366.1 | 347.4 | 338.4 | 329.8 | 321.5 | 313.4 | 305.6 | 298.1 |
| 3 | 7359. | 590.6 | 273.3 | 189.2 | 152.7 | 132.8 | 120.3 | 111.8 | 105.6 | 101.0 | 94.42 | 88.34 | 82.70 | 80.03 | 77.46 | 74.97 | 72.58 | 70.28 | 68.05 |
| 4 | 4761. | 346.4 | 152.2 | 102.1 | 80.59 | 68.94 | 61.71 | 56.80 | 53.26 | 50.59 | 46.45 | 43.38 | 40.18 | 38.67 | 37.22 | 35.82 | 34.47 | 33.18 | 31.93 |
| 5 | 3742. | 256.1 | 108.7 | 71.27 | 55.39 | 46.82 | 41.53 | 37.94 | 35.37 | 33.43 | 30.71 | 28.21 | 25.90 | 24.81 | 23.76 | 22.76 | 21.79 | 20.87 | 19.97 |
| 6 | 3316. | 211.2 | 87.43 | 56.38 | 43.29 | 36.27 | 31.93 | 29.01 | 26.91 | 25.33 | 23.12 | 21.09 | 19.22 | 18.34 | 17.50 | 16.69 | 15.91 | 15.16 | 14.44 |
| 7 | 2999. | 184.8 | 75.11 | 47.82 | 36.38 | 30.25 | 26.48 | 23.94 | 22.12 | 20.76 | 18.76 | 17.09 | 15.47 | 14.72 | 13.99 | 13.29 | 12.62 | 11.97 | 11.35 |
| 8 | 2890. | 167.5 | 67.14 | 42.32 | 31.95 | 26.42 | 23.02 | 20.73 | 19.09 | 17.86 | 16.14 | 14.56 | 13.11 | 12.43 | 11.78 | 11.15 | 10.55 | 9.973 | 9.418 |
| 9 | 2540. | 155.6 | 61.61 | 38.52 | 28.91 | 23.78 | 20.64 | 18.52 | 17.01 | 15.88 | 14.29 | 12.84 | 11.50 | 10.88 | 10.28 | 9.700 | 9.146 | 8.614 | 8.103 |
| 10 | 2429. | 146.6 | 57.57 | 35.75 | 26.69 | 21.87 | 18.91 | 16.93 | 15.51 | 14.44 | 12.96 | 11.59 | 10.34 | 9.756 | 9.193 | 8.652 | 8.134 | 7.635 | 7.156 |
| 12 | 2276. | 134.5 | 52.05 | 31.99 | 23.69 | 19.28 | 16.58 | 14.78 | 13.48 | 12.51 | 11.16 | 9.926 | 8.788 | 8.254 | 7.742 | 7.251 | 6.779 | 6.325 | 5.889 |
| 15 | 2136. | 123.5 | 47.14 | 28.66 | 21.04 | 17.01 | 14.54 | 12.89 | 11.71 | 10.82 | 9.594 | 8.406 | 7.439 | 6.943 | 6.476 | 6.027 | 5.596 | 5.181 | 4.780 |
| 20 | 2008. | 113.7 | 42.77 | 25.71 | 18.70 | 15.00 | 12.74 | 11.23 | 10.15 | 9.343 | 8.219 | 7.187 | 6.240 | 5.795 | 5.367 | 4.955 | 4.558 | 4.175 | 3.802 |
| 24 | 1948. | 109.2 | 40.76 | 24.36 | 17.63 | 14.09 | 11.92 | 10.48 | 9.444 | 8.672 | 7.596 | 6.649 | 5.702 | 5.275 | 4.864 | 4.468 | 4.086 | 3.715 | 3.354 |
| 30 | 1891. | 104.9 | 38.86 | 23.09 | 16.63 | 13.23 | 11.16 | 9.770 | 8.782 | 8.042 | 7.012 | 6.067 | 5.197 | 4.787 | 4.392 | 4.011 | 3.641 | 3.281 | 2.927 |
| 40 | 1837. | 100.8 | 37.06 | 21.89 | 15.68 | 12.42 | 10.43 | 9.107 | 8.160 | 7.452 | 6.465 | 5.558 | 4.772 | 4.329 | 3.948 | 3.579 | 3.220 | 2.867 | 2.516 |
| 60 | 1785. | 96.95 | 35.37 | 20.76 | 14.80 | 11.66 | 9.758 | 8.485 | 7.577 | 6.898 | 5.951 | 5.081 | 4.278 | 3.997 | 3.529 | 3.170 | 2.817 | 2.467 | 2.110 |
| 120 | 1735. | 93.27 | 33.76 | 19.69 | 13.96 | 10.95 | 9.122 | 7.901 | 7.030 | 6.378 | 5.469 | 4.633 | 3.855 | 3.490 | 3.131 | 2.779 | 2.429 | 2.072 | 1.688 |
| ∞ | 1687. | 89.78 | 32.24 | 18.68 | 13.17 | 10.28 | 8.524 | 7.352 | 6.516 | 5.890 | 5.017 | 4.211 | 3.462 | 3.104 | 2.752 | 2.403 | 2.046 | 1.661 | 1.000 |

For explanation see the panel of this table for β = 0.005.

TABLE 8.4.—$\phi(0.01, \beta, n_1, n_2)$.—(Continued)

$$\beta = 0.10$$

| $n_2 \backslash n_1$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 15 | 20 | 24 | 30 | 40 | 60 | 120 | ∞ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 161500. | 42630. | 29030. | 25560. | 23400. | 22120. | 21280. | 20880. | 20240. | 19890. | 19400. | 18920. | 18470. | 18250. | 18040. | 17830. | 17620. | 17420. | 17290. |
| 2 | 4876. | 891.0 | 541.7 | 429.2 | 375.3 | 344.0 | 323.6 | 309.4 | 298.8 | 290.7 | 279.0 | 268.0 | 257.5 | 252.5 | 247.5 | 242.8 | 238.1 | 233.5 | 229.1 |
| 3 | 1828. | 282.3 | 158.8 | 120.3 | 102.2 | 91.79 | 85.07 | 80.37 | 76.92 | 74.27 | 70.48 | 66.91 | 63.52 | 61.90 | 60.33 | 58.79 | 57.30 | 55.85 | 54.44 |
| 4 | 1184. | 166.4 | 89.19 | 65.62 | 54.64 | 48.37 | 44.34 | 41.53 | 39.47 | 37.90 | 35.65 | 33.53 | 31.53 | 30.57 | 29.64 | 28.74 | 27.86 | 27.01 | 26.18 |
| 5 | 980.6 | 123.4 | 64.03 | 46.14 | 37.87 | 33.16 | 30.15 | 28.05 | 26.52 | 25.34 | 23.67 | 22.10 | 20.62 | 19.91 | 19.22 | 18.55 | 17.90 | 17.28 | 16.66 |
| 6 | 800.0 | 101.9 | 51.68 | 36.68 | 29.78 | 25.86 | 23.35 | 21.62 | 20.35 | 19.38 | 17.99 | 16.69 | 15.47 | 14.88 | 14.31 | 13.76 | 13.23 | 12.71 | 12.21 |
| 7 | 731.4 | 90.25 | 44.51 | 31.22 | 25.13 | 21.68 | 19.47 | 17.95 | 16.83 | 15.98 | 14.77 | 13.63 | 12.56 | 12.04 | 11.55 | 11.06 | 10.60 | 10.14 | 9.698 |
| 8 | 669.2 | 81.01 | 39.87 | 27.71 | 22.15 | 19.00 | 17.00 | 15.61 | 14.60 | 13.82 | 12.72 | 11.68 | 10.71 | 10.25 | 9.794 | 9.356 | 8.930 | 8.517 | 8.115 |
| 9 | 632.2 | 75.25 | 36.64 | 25.28 | 20.09 | 17.16 | 15.29 | 14.00 | 13.06 | 12.34 | 11.31 | 10.35 | 9.447 | 9.015 | 8.595 | 8.188 | 7.792 | 7.407 | 7.033 |
| 10 | 604.6 | 70.99 | 34.27 | 23.50 | 18.59 | 15.82 | 14.05 | 12.83 | 11.94 | 11.26 | 10.30 | 9.387 | 8.532 | 8.124 | 7.727 | 7.342 | 6.968 | 6.604 | 6.249 |
| 12 | 566.4 | 65.17 | 31.05 | 21.08 | 16.55 | 14.00 | 12.38 | 11.26 | 10.44 | 9.813 | 8.923 | 8.068 | 7.302 | 6.925 | 6.560 | 6.205 | 5.860 | 5.523 | 5.195 |
| 15 | 531.6 | 59.93 | 28.17 | 18.93 | 14.75 | 12.40 | 10.90 | 9.888 | 9.112 | 8.536 | 7.717 | 6.946 | 6.221 | 5.874 | 5.536 | 5.206 | 4.886 | 4.572 | 4.266 |
| 20 | 499.8 | 55.22 | 25.60 | 17.03 | 13.16 | 10.98 | 9.597 | 8.642 | 7.945 | 7.412 | 6.655 | 5.942 | 5.270 | 4.947 | 4.633 | 4.326 | 4.025 | 3.730 | 3.440 |
| 24 | 485.0 | 53.05 | 24.42 | 16.16 | 12.43 | 10.33 | 9.003 | 8.085 | 7.413 | 6.901 | 6.172 | 5.486 | 4.837 | 4.526 | 4.221 | 3.923 | 3.631 | 3.343 | 3.058 |
| 30 | 470.9 | 50.98 | 23.31 | 15.34 | 11.74 | 9.726 | 8.445 | 7.560 | 6.914 | 6.421 | 5.719 | 5.057 | 4.430 | 4.128 | 3.833 | 3.543 | 3.258 | 2.975 | 2.692 |
| 40 | 457.3 | 49.02 | 22.25 | 14.56 | 11.09 | 9.153 | 7.919 | 7.068 | 6.445 | 5.970 | 5.293 | 4.653 | 4.047 | 3.754 | 3.466 | 3.183 | 2.902 | 2.622 | 2.337 |
| 60 | 444.4 | 47.16 | 21.25 | 13.83 | 10.48 | 8.614 | 7.424 | 6.604 | 6.004 | 5.546 | 4.892 | 4.274 | 3.685 | 3.400 | 3.119 | 2.840 | 2.562 | 2.279 | 1.985 |
| 120 | 432.0 | 45.39 | 20.31 | 13.14 | 9.910 | 8.106 | 6.959 | 6.168 | 5.589 | 5.146 | 4.514 | 3.916 | 3.343 | 3.064 | 2.788 | 2.512 | 2.231 | 1.939 | 1.613 |
| ∞ | 420.2 | 43.71 | 19.41 | 12.48 | 9.369 | 7.628 | 6.521 | 5.757 | 5.198 | 4.770 | 4.159 | 3.578 | 3.019 | 2.745 | 2.471 | 2.192 | 1.902 | 1.580 | 1.000 |

For explanation see the panel of this table for $\beta = 0.005$.

TABLE 8.4    COMPARING TWO STANDARD DEVIATIONS    303

β = 0.25

| n₁ \\ n₂ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 15 | 20 | 24 | 30 | 40 | 60 | 120 | ∞ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 23650. | 12860. | 10940. | 10170. | 9755. | 9500. | 9326. | 9202. | 9107. | 9032. | 8923. | 8818. | 8715. | 8665. | 8615. | 8566. | 8518. | 8471. | 8424. |
| 2 | 738.8 | 297.0 | 226.1 | 198.5 | 184.0 | 175.0 | 169.0 | 164.7 | 161.4 | 158.8 | 155.0 | 151.4 | 147.9 | 146.2 | 144.5 | 142.8 | 141.1 | 139.5 | 137.9 |
| 3 | 270.7 | 97.18 | 69.39 | 58.76 | 53.21 | 49.80 | 47.51 | 45.86 | 44.61 | 43.64 | 42.23 | 40.85 | 39.52 | 38.87 | 38.24 | 37.61 | 36.99 | 36.38 | 35.78 |
| 4 | 181.9 | 58.18 | 39.90 | 32.98 | 29.38 | 27.18 | 25.69 | 24.63 | 23.83 | 23.20 | 22.29 | 21.40 | 20.54 | 20.12 | 19.71 | 19.30 | 18.91 | 18.51 | 18.12 |
| 5 | 143.4 | 43.54 | 29.06 | 23.61 | 20.78 | 19.05 | 17.89 | 17.05 | 16.43 | 15.93 | 15.22 | 14.52 | 13.85 | 13.52 | 13.20 | 12.88 | 12.57 | 12.26 | 11.95 |
| 6 | 123.5 | 36.18 | 23.68 | 19.00 | 16.57 | 15.09 | 14.09 | 13.37 | 12.83 | 12.41 | 11.80 | 11.20 | 10.62 | 10.34 | 10.06 | 9.790 | 9.520 | 9.253 | 8.901 |
| 7 | 111.5 | 31.84 | 20.54 | 16.57 | 14.13 | 12.79 | 11.90 | 11.25 | 10.76 | 10.39 | 9.831 | 9.283 | 8.773 | 8.517 | 8.266 | 8.018 | 7.774 | 7.532 | 7.294 |
| 8 | 103.5 | 29.00 | 18.49 | 14.58 | 12.55 | 11.31 | 10.48 | 9.885 | 9.434 | 9.083 | 8.568 | 8.059 | 7.586 | 7.348 | 7.114 | 6.883 | 6.655 | 6.429 | 6.207 |
| 9 | 97.83 | 27.00 | 17.07 | 13.37 | 11.45 | 10.29 | 9.503 | 8.939 | 8.513 | 8.181 | 7.695 | 7.223 | 6.764 | 6.539 | 6.317 | 6.098 | 5.881 | 5.667 | 5.455 |
| 10 | 93.61 | 25.53 | 16.02 | 12.48 | 10.65 | 9.537 | 8.787 | 8.247 | 7.840 | 7.523 | 7.057 | 6.604 | 6.165 | 5.949 | 5.736 | 5.526 | 5.317 | 5.110 | 4.905 |
| 12 | 87.76 | 23.50 | 14.58 | 11.27 | 9.560 | 8.517 | 7.814 | 7.309 | 6.927 | 6.629 | 6.192 | 5.764 | 5.358 | 5.149 | 4.948 | 4.748 | 4.550 | 4.353 | 4.158 |
| 15 | 82.43 | 21.68 | 13.30 | 10.19 | 8.588 | 7.600 | 6.950 | 6.475 | 6.117 | 5.836 | 5.426 | 5.094 | 4.632 | 4.438 | 4.247 | 4.056 | 3.867 | 3.678 | 3.499 |
| 20 | 77.87 | 20.04 | 12.15 | 9.228 | 7.721 | 6.802 | 6.181 | 5.734 | 5.396 | 5.131 | 4.743 | 4.362 | 3.989 | 3.805 | 3.621 | 3.438 | 3.255 | 3.070 | 2.885 |
| 24 | 75.30 | 19.28 | 11.63 | 8.786 | 7.324 | 6.431 | 5.829 | 5.395 | 5.066 | 4.809 | 4.430 | 4.069 | 3.695 | 3.514 | 3.333 | 3.153 | 2.971 | 2.788 | 2.601 |
| 30 | 73.13 | 18.56 | 11.12 | 8.307 | 6.948 | 6.082 | 5.497 | 5.075 | 4.755 | 4.504 | 4.135 | 3.773 | 3.416 | 3.237 | 3.060 | 2.881 | 2.700 | 2.516 | 2.327 |
| 40 | 71.06 | 17.87 | 10.64 | 7.971 | 6.593 | 5.752 | 5.183 | 4.772 | 4.461 | 4.216 | 3.856 | 3.502 | 3.151 | 2.975 | 2.799 | 2.621 | 2.440 | 2.253 | 2.058 |
| 60 | 69.07 | 17.22 | 10.19 | 7.596 | 6.258 | 5.440 | 4.887 | 4.487 | 4.184 | 3.945 | 3.593 | 3.245 | 2.909 | 2.736 | 2.550 | 2.371 | 2.187 | 1.995 | 1.787 |
| 120 | 67.17 | 16.60 | 9.763 | 7.242 | 5.940 | 5.145 | 4.607 | 4.217 | 3.921 | 3.688 | 3.343 | 3.002 | 2.660 | 2.487 | 2.311 | 2.129 | 1.940 | 1.734 | 1.496 |
| ∞ | 65.35 | 16.01 | 9.356 | 6.906 | 5.641 | 4.867 | 4.342 | 3.962 | 3.673 | 3.445 | 3.107 | 2.771 | 2.431 | 2.258 | 2.079 | 1.892 | 1.690 | 1.455 | 1.000 |

For explanation see the panel of this table for β = 0.005.

TABLE 8.4.—$\phi(0.01, \beta, n_1, n_2)$.—(Continued)

$\beta = 0.50$

| $n_2 \backslash n_1$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 15 | 20 | 24 | 30 | 40 | 60 | 120 | ∞ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 4052. | 3333. | 3161. | 3086. | 3044. | 3017. | 2998. | 2985. | 2974. | 2966. | 2954. | 2942. | 2930. | 2924. | 2918. | 2913. | 2907. | 2902. | 2896. |
| 2 | 147.8 | 99.00 | 87.38 | 82.22 | 79.32 | 77.46 | 76.16 | 75.21 | 74.48 | 73.90 | 73.05 | 72.20 | 71.38 | 70.97 | 70.56 | 70.16 | 69.76 | 69.36 | 68.97 |
| 3 | 58.31 | 34.98 | 29.46 | 27.00 | 25.61 | 24.72 | 24.10 | 23.64 | 23.29 | 23.01 | 22.60 | 22.19 | 21.78 | 21.59 | 21.39 | 21.19 | 20.99 | 20.80 | 20.60 |
| 4 | 38.64 | 21.73 | 17.75 | 15.98 | 14.97 | 14.32 | 13.87 | 13.54 | 13.28 | 13.07 | 12.77 | 12.47 | 12.17 | 12.03 | 11.88 | 11.73 | 11.59 | 11.44 | 11.30 |
| 5 | 30.79 | 16.62 | 13.29 | 11.81 | 10.97 | 10.42 | 10.04 | 9.757 | 9.540 | 9.367 | 9.109 | 8.854 | 8.601 | 8.475 | 8.349 | 8.224 | 8.100 | 7.975 | 7.851 |
| 6 | 26.70 | 14.01 | 11.04 | 9.712 | 8.956 | 8.466 | 8.123 | 7.867 | 7.671 | 7.515 | 7.281 | 7.050 | 6.820 | 6.705 | 6.590 | 6.476 | 6.361 | 6.247 | 6.133 |
| 7 | 24.21 | 12.45 | 9.704 | 8.472 | 7.769 | 7.313 | 6.993 | 6.755 | 6.572 | 6.425 | 6.207 | 5.989 | 5.772 | 5.664 | 5.556 | 5.448 | 5.339 | 5.230 | 5.121 |
| 8 | 22.56 | 11.43 | 8.826 | 7.660 | 6.993 | 6.560 | 6.255 | 6.029 | 5.854 | 5.714 | 5.505 | 5.297 | 5.089 | 4.985 | 4.880 | 4.776 | 4.671 | 4.566 | 4.460 |
| 9 | 21.39 | 10.70 | 8.210 | 7.090 | 6.449 | 6.033 | 5.738 | 5.520 | 5.351 | 5.216 | 5.014 | 4.812 | 4.610 | 4.509 | 4.407 | 4.305 | 4.203 | 4.100 | 3.996 |
| 10 | 20.51 | 10.17 | 7.753 | 6.669 | 6.048 | 5.643 | 5.358 | 5.145 | 4.981 | 4.849 | 4.652 | 4.455 | 4.257 | 4.157 | 4.057 | 3.957 | 3.856 | 3.754 | 3.652 |
| 12 | 19.29 | 9.427 | 7.126 | 6.091 | 5.497 | 5.110 | 4.836 | 4.631 | 4.473 | 4.346 | 4.155 | 3.964 | 3.771 | 3.674 | 3.577 | 3.478 | 3.379 | 3.278 | 3.176 |
| 15 | 18.17 | 8.757 | 6.561 | 5.571 | 5.002 | 4.630 | 4.366 | 4.169 | 4.016 | 3.893 | 3.708 | 3.522 | 3.334 | 3.238 | 3.142 | 3.045 | 2.945 | 2.845 | 2.742 |
| 20 | 17.16 | 8.149 | 6.050 | 5.103 | 4.557 | 4.198 | 3.944 | 3.754 | 3.605 | 3.486 | 3.306 | 3.123 | 2.938 | 2.843 | 2.747 | 2.649 | 2.550 | 2.447 | 2.341 |
| 24 | 16.68 | 7.867 | 5.814 | 4.886 | 4.351 | 3.999 | 3.749 | 3.562 | 3.415 | 3.297 | 3.119 | 2.938 | 2.754 | 2.659 | 2.563 | 2.464 | 2.363 | 2.259 | 2.150 |
| 30 | 16.22 | 7.598 | 5.589 | 4.680 | 4.155 | 3.810 | 3.564 | 3.379 | 3.235 | 3.118 | 2.942 | 2.762 | 2.578 | 2.483 | 2.386 | 2.286 | 2.183 | 2.075 | 1.962 |
| 40 | 15.79 | 7.342 | 5.375 | 4.485 | 3.970 | 3.630 | 3.388 | 3.206 | 3.063 | 2.948 | 2.773 | 2.594 | 2.409 | 2.314 | 2.216 | 2.114 | 2.008 | 1.896 | 1.775 |
| 60 | 15.37 | 7.098 | 5.172 | 4.299 | 3.793 | 3.460 | 3.221 | 3.041 | 2.900 | 2.786 | 2.612 | 2.433 | 2.248 | 2.151 | 2.051 | 1.947 | 1.836 | 1.717 | 1.563 |
| 120 | 14.97 | 6.866 | 4.979 | 4.123 | 3.626 | 3.298 | 3.062 | 2.894 | 2.745 | 2.631 | 2.458 | 2.280 | 2.093 | 1.994 | 1.892 | 1.783 | 1.665 | 1.533 | 1.373 |
| ∞ | 14.58 | 6.644 | 4.795 | 3.945 | 3.467 | 3.144 | 2.911 | 2.736 | 2.597 | 2.484 | 2.312 | 2.132 | 1.943 | 1.842 | 1.735 | 1.619 | 1.490 | 1.332 | 1.000 |

For explanation see the panel of this table for $\beta = 0.005$.

TABLE 8.4   COMPARING TWO STANDARD DEVIATIONS   305

β = 0.75

| $n_1$ \ $n_2$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 15 | 20 | 24 | 30 | 40 | 60 | 120 | ∞ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 605.2 | 666.6 | 658.9 | 655.5 | 653.5 | 652.2 | 651.3 | 650.7 | 650.2 | 649.8 | 649.2 | 648.6 | 648.0 | 647.7 | 647.4 | 647.2 | 646.9 | 646.6 | 646.3 |
| 2 | 38.31 | 33.00 | 31.45 | 30.71 | 30.28 | 29.99 | 29.79 | 29.64 | 29.53 | 29.43 | 29.30 | 29.16 | 29.03 | 28.96 | 28.89 | 28.82 | 28.76 | 28.69 | 28.62 |
| 3 | 16.86 | 13.52 | 12.51 | 12.01 | 11.72 | 11.52 | 11.39 | 11.28 | 11.20 | 11.14 | 11.04 | 10.94 | 10.85 | 10.80 | 10.75 | 10.70 | 10.66 | 10.61 | 10.56 |
| 4 | 11.73 | 9.000 | 8.157 | 7.740 | 7.490 | 7.323 | 7.203 | 7.113 | 7.043 | 6.987 | 6.902 | 6.816 | 6.731 | 6.688 | 6.645 | 6.602 | 6.558 | 6.515 | 6.471 |
| 5 | 9.606 | 7.164 | 6.400 | 6.019 | 5.788 | 5.633 | 5.522 | 5.437 | 5.371 | 5.318 | 5.238 | 5.157 | 5.076 | 5.035 | 4.993 | 4.952 | 4.910 | 4.868 | 4.825 |
| 6 | 8.477 | 6.200 | 5.481 | 5.119 | 4.899 | 4.751 | 4.643 | 4.562 | 4.498 | 4.447 | 4.369 | 4.290 | 4.210 | 4.169 | 4.128 | 4.087 | 4.046 | 4.004 | 3.961 |
| 7 | 7.784 | 5.612 | 4.922 | 4.573 | 4.360 | 4.216 | 4.111 | 4.031 | 3.968 | 3.918 | 3.841 | 3.763 | 3.683 | 3.643 | 3.602 | 3.561 | 3.519 | 3.477 | 3.434 |
| 8 | 7.319 | 5.220 | 4.550 | 4.210 | 4.001 | 3.859 | 3.756 | 3.677 | 3.615 | 3.565 | 3.490 | 3.411 | 3.351 | 3.291 | 3.250 | 3.208 | 3.166 | 3.123 | 3.080 |
| 9 | 6.984 | 4.941 | 4.286 | 3.951 | 3.746 | 3.606 | 3.503 | 3.425 | 3.364 | 3.314 | 3.238 | 3.160 | 3.080 | 3.039 | 2.998 | 2.956 | 2.913 | 2.870 | 2.825 |
| 10 | 6.734 | 4.732 | 4.088 | 3.758 | 3.555 | 3.416 | 3.315 | 3.237 | 3.176 | 3.126 | 3.050 | 2.974 | 2.892 | 2.851 | 2.809 | 2.767 | 2.723 | 2.679 | 2.634 |
| 12 | 6.385 | 4.442 | 3.814 | 3.491 | 3.291 | 3.154 | 3.053 | 2.976 | 2.915 | 2.865 | 2.788 | 2.713 | 2.598 | 2.537 | 2.545 | 2.501 | 2.456 | 2.410 | 2.363 |
| 15 | 6.063 | 4.176 | 3.563 | 3.247 | 3.050 | 2.914 | 2.814 | 2.737 | 2.676 | 2.626 | 2.549 | 2.474 | 2.408 | 2.344 | 2.300 | 2.255 | 2.209 | 2.161 | 2.111 |
| 20 | 5.768 | 3.933 | 3.335 | 3.024 | 2.829 | 2.695 | 2.595 | 2.518 | 2.457 | 2.407 | 2.329 | 2.246 | 2.165 | 2.119 | 2.073 | 2.026 | 1.977 | 1.925 | 1.871 |
| 24 | 5.629 | 3.830 | 3.228 | 2.920 | 2.727 | 2.593 | 2.493 | 2.416 | 2.355 | 2.304 | 2.226 | 2.144 | 2.057 | 2.012 | 1.965 | 1.917 | 1.865 | 1.811 | 1.754 |
| 30 | 5.496 | 3.711 | 3.126 | 2.821 | 2.628 | 2.495 | 2.395 | 2.318 | 2.256 | 2.206 | 2.127 | 2.044 | 1.956 | 1.909 | 1.861 | 1.810 | 1.756 | 1.699 | 1.637 |
| 40 | 5.368 | 3.607 | 3.029 | 2.726 | 2.535 | 2.401 | 2.302 | 2.224 | 2.162 | 2.111 | 2.031 | 1.947 | 1.857 | 1.809 | 1.759 | 1.705 | 1.649 | 1.587 | 1.519 |
| 60 | 5.245 | 3.506 | 2.936 | 2.635 | 2.445 | 2.312 | 2.212 | 2.135 | 2.072 | 2.020 | 1.939 | 1.854 | 1.761 | 1.711 | 1.659 | 1.603 | 1.542 | 1.474 | 1.395 |
| 120 | 5.127 | 3.413 | 2.847 | 2.548 | 2.359 | 2.226 | 2.127 | 2.049 | 1.986 | 1.933 | 1.851 | 1.763 | 1.667 | 1.616 | 1.560 | 1.500 | 1.433 | 1.355 | 1.256 |
| ∞ | 5.014 | 3.322 | 2.762 | 2.465 | 2.277 | 2.144 | 2.044 | 1.966 | 1.902 | 1.849 | 1.766 | 1.676 | 1.577 | 1.522 | 1.462 | 1.396 | 1.319 | 1.222 | 1.000 |

For explanation see the panel of this table for β = 0.005.

TABLE 8.4.—$\phi(0.01,\beta,n_1,n_2)$.—(Continued)

$\beta = 0.90$

| $n_2$ \ $n_1$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 15 | 20 | 24 | 30 | 40 | 60 | 120 | ∞ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 101.7 | 101.0 | 100.8 | 100.7 | 100.7 | 100.7 | 100.6 | 100.6 | 100.6 | 100.6 | 100.6 | 100.6 | 100.6 | 100.6 | 100.5 | 100.5 | 100.5 | 100.5 | 100.5 |
| 2 | 11.55 | 11.00 | 10.82 | 10.74 | 10.69 | 10.65 | 10.63 | 10.61 | 10.60 | 10.58 | 10.57 | 10.55 | 10.53 | 10.53 | 10.52 | 10.51 | 10.50 | 10.49 | 10.48 |
| 3 | 6.160 | 5.642 | 5.464 | 5.374 | 5.319 | 5.281 | 5.255 | 5.234 | 5.219 | 5.206 | 5.187 | 5.167 | 5.148 | 5.138 | 5.129 | 5.119 | 5.109 | 5.099 | 5.089 |
| 4 | 4.664 | 4.162 | 3.983 | 3.890 | 3.832 | 3.792 | 3.764 | 3.742 | 3.725 | 3.711 | 3.690 | 3.670 | 3.647 | 3.636 | 3.625 | 3.614 | 3.602 | 3.591 | 3.580 |
| 5 | 4.004 | 3.512 | 3.332 | 3.236 | 3.176 | 3.135 | 3.105 | 3.081 | 3.063 | 3.048 | 3.026 | 3.003 | 2.979 | 2.967 | 2.955 | 2.943 | 2.930 | 2.918 | 2.905 |
| 6 | 3.640 | 3.155 | 2.974 | 2.876 | 2.814 | 2.772 | 2.740 | 2.716 | 2.697 | 2.681 | 2.657 | 2.633 | 2.608 | 2.595 | 2.582 | 2.568 | 2.555 | 2.541 | 2.527 |
| 7 | 3.413 | 2.931 | 2.749 | 2.650 | 2.587 | 2.543 | 2.511 | 2.486 | 2.466 | 2.450 | 2.425 | 2.399 | 2.372 | 2.359 | 2.345 | 2.331 | 2.316 | 2.302 | 2.287 |
| 8 | 3.256 | 2.778 | 2.596 | 2.496 | 2.432 | 2.388 | 2.354 | 2.328 | 2.308 | 2.291 | 2.265 | 2.238 | 2.210 | 2.196 | 2.181 | 2.166 | 2.151 | 2.135 | 2.119 |
| 9 | 3.143 | 2.668 | 2.486 | 2.385 | 2.320 | 2.274 | 2.240 | 2.214 | 2.193 | 2.175 | 2.149 | 2.121 | 2.092 | 2.077 | 2.062 | 2.046 | 2.030 | 2.013 | 1.996 |
| 10 | 3.058 | 2.585 | 2.402 | 2.301 | 2.235 | 2.189 | 2.154 | 2.127 | 2.106 | 2.088 | 2.060 | 2.032 | 2.002 | 1.986 | 1.970 | 1.954 | 1.937 | 1.920 | 1.902 |
| 12 | 2.937 | 2.468 | 2.285 | 2.182 | 2.115 | 2.068 | 2.032 | 2.005 | 1.982 | 1.964 | 1.935 | 1.905 | 1.873 | 1.857 | 1.840 | 1.822 | 1.804 | 1.785 | 1.765 |
| 15 | 2.825 | 2.359 | 2.176 | 2.072 | 2.004 | 1.956 | 1.919 | 1.890 | 1.867 | 1.848 | 1.818 | 1.786 | 1.752 | 1.735 | 1.716 | 1.697 | 1.677 | 1.656 | 1.634 |
| 20 | 2.722 | 2.259 | 2.075 | 1.970 | 1.901 | 1.851 | 1.813 | 1.784 | 1.759 | 1.739 | 1.707 | 1.674 | 1.638 | 1.618 | 1.598 | 1.577 | 1.555 | 1.532 | 1.506 |
| 24 | 2.673 | 2.212 | 2.027 | 1.922 | 1.852 | 1.802 | 1.763 | 1.733 | 1.708 | 1.687 | 1.655 | 1.620 | 1.582 | 1.562 | 1.541 | 1.519 | 1.495 | 1.470 | 1.442 |
| 30 | 2.625 | 2.166 | 1.981 | 1.876 | 1.805 | 1.754 | 1.715 | 1.684 | 1.658 | 1.637 | 1.604 | 1.568 | 1.529 | 1.508 | 1.485 | 1.461 | 1.436 | 1.408 | 1.378 |
| 40 | 2.580 | 2.122 | 1.937 | 1.831 | 1.760 | 1.708 | 1.668 | 1.637 | 1.611 | 1.589 | 1.554 | 1.517 | 1.476 | 1.454 | 1.430 | 1.404 | 1.376 | 1.346 | 1.311 |
| 60 | 2.535 | 2.080 | 1.895 | 1.788 | 1.716 | 1.664 | 1.623 | 1.591 | 1.564 | 1.542 | 1.506 | 1.467 | 1.424 | 1.400 | 1.375 | 1.347 | 1.316 | 1.281 | 1.239 |
| 120 | 2.493 | 2.039 | 1.854 | 1.747 | 1.674 | 1.621 | 1.580 | 1.546 | 1.519 | 1.496 | 1.459 | 1.418 | 1.373 | 1.347 | 1.320 | 1.289 | 1.254 | 1.212 | 1.158 |
| ∞ | 2.452 | 2.000 | 1.815 | 1.707 | 1.633 | 1.579 | 1.537 | 1.504 | 1.476 | 1.452 | 1.413 | 1.371 | 1.322 | 1.295 | 1.264 | 1.229 | 1.188 | 1.133 | 1.000 |

For explanation see the panel of this table for $\beta = 0.005$.

TABLE 8.4   COMPARING TWO STANDARD DEVIATIONS   307

$$\beta = 0.95$$

| $n_2 \backslash n_1$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 15 | 20 | 24 | 30 | 40 | 60 | 120 | ∞ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 25.10 | 25.06 | 25.05 | 25.04 | 25.04 | 25.04 | 25.04 | 25.04 | 25.04 | 25.04 | 25.04 | 25.03 | 25.03 | 25.03 | 25.03 | 25.03 | 25.03 | 25.03 | 25.03 |
| 2 | 5.321 | 5.211 | 5.175 | 5.157 | 5.146 | 5.139 | 5.134 | 5.130 | 5.127 | 5.125 | 5.121 | 5.118 | 5.114 | 5.112 | 5.111 | 5.109 | 5.107 | 5.106 | 5.104 |
| 3 | 3.368 | 3.226 | 3.175 | 3.149 | 3.133 | 3.122 | 3.114 | 3.108 | 3.103 | 3.099 | 3.094 | 3.088 | 3.082 | 3.079 | 3.076 | 3.073 | 3.070 | 3.067 | 3.064 |
| 4 | 2.750 | 2.592 | 2.533 | 2.501 | 2.481 | 2.467 | 2.457 | 2.450 | 2.444 | 2.439 | 2.431 | 2.424 | 2.416 | 2.412 | 2.408 | 2.404 | 2.400 | 2.396 | 2.392 |
| 5 | 2.460 | 2.294 | 2.229 | 2.194 | 2.172 | 2.156 | 2.144 | 2.135 | 2.128 | 2.123 | 2.114 | 2.105 | 2.096 | 2.091 | 2.086 | 2.081 | 2.077 | 2.072 | 2.067 |
| 6 | 2.296 | 2.124 | 2.056 | 2.018 | 1.993 | 1.976 | 1.964 | 1.954 | 1.946 | 1.939 | 1.930 | 1.919 | 1.909 | 1.904 | 1.898 | 1.893 | 1.887 | 1.881 | 1.875 |
| 7 | 2.190 | 2.015 | 1.944 | 1.904 | 1.878 | 1.860 | 1.847 | 1.836 | 1.827 | 1.820 | 1.810 | 1.799 | 1.787 | 1.781 | 1.775 | 1.769 | 1.762 | 1.756 | 1.749 |
| 8 | 2.117 | 1.940 | 1.867 | 1.826 | 1.798 | 1.779 | 1.765 | 1.754 | 1.745 | 1.737 | 1.726 | 1.714 | 1.701 | 1.695 | 1.688 | 1.681 | 1.674 | 1.667 | 1.660 |
| 9 | 2.064 | 1.885 | 1.810 | 1.768 | 1.740 | 1.720 | 1.705 | 1.693 | 1.683 | 1.675 | 1.663 | 1.651 | 1.637 | 1.630 | 1.623 | 1.616 | 1.608 | 1.601 | 1.593 |
| 10 | 2.023 | 1.842 | 1.767 | 1.723 | 1.695 | 1.674 | 1.658 | 1.646 | 1.636 | 1.628 | 1.615 | 1.604 | 1.588 | 1.581 | 1.573 | 1.565 | 1.557 | 1.549 | 1.540 |
| 12 | 1.965 | 1.783 | 1.705 | 1.660 | 1.631 | 1.609 | 1.592 | 1.580 | 1.569 | 1.560 | 1.547 | 1.534 | 1.517 | 1.509 | 1.501 | 1.492 | 1.483 | 1.473 | 1.464 |
| 15 | 1.911 | 1.727 | 1.648 | 1.601 | 1.570 | 1.548 | 1.530 | 1.516 | 1.505 | 1.496 | 1.481 | 1.468 | 1.449 | 1.440 | 1.431 | 1.421 | 1.411 | 1.400 | 1.389 |
| 20 | 1.861 | 1.675 | 1.594 | 1.546 | 1.513 | 1.490 | 1.471 | 1.457 | 1.445 | 1.435 | 1.419 | 1.402 | 1.383 | 1.373 | 1.363 | 1.352 | 1.340 | 1.327 | 1.314 |
| 24 | 1.836 | 1.650 | 1.568 | 1.519 | 1.486 | 1.462 | 1.443 | 1.428 | 1.416 | 1.405 | 1.388 | 1.371 | 1.351 | 1.340 | 1.329 | 1.317 | 1.305 | 1.291 | 1.276 |
| 30 | 1.813 | 1.626 | 1.543 | 1.494 | 1.460 | 1.435 | 1.416 | 1.400 | 1.387 | 1.376 | 1.359 | 1.346 | 1.319 | 1.308 | 1.296 | 1.283 | 1.269 | 1.254 | 1.237 |
| 40 | 1.791 | 1.602 | 1.519 | 1.469 | 1.434 | 1.409 | 1.389 | 1.373 | 1.360 | 1.348 | 1.330 | 1.310 | 1.288 | 1.276 | 1.263 | 1.249 | 1.233 | 1.216 | 1.196 |
| 60 | 1.769 | 1.580 | 1.496 | 1.445 | 1.410 | 1.384 | 1.363 | 1.346 | 1.333 | 1.321 | 1.302 | 1.281 | 1.257 | 1.244 | 1.230 | 1.214 | 1.197 | 1.177 | 1.152 |
| 120 | 1.748 | 1.558 | 1.474 | 1.422 | 1.386 | 1.359 | 1.338 | 1.321 | 1.306 | 1.294 | 1.274 | 1.252 | 1.227 | 1.212 | 1.197 | 1.179 | 1.159 | 1.134 | 1.101 |
| ∞ | 1.727 | 1.537 | 1.452 | 1.399 | 1.363 | 1.335 | 1.313 | 1.296 | 1.281 | 1.268 | 1.247 | 1.223 | 1.196 | 1.180 | 1.163 | 1.142 | 1.118 | 1.064 | 1.000 |

For explanation see the panel of this table for $\beta = 0.005$.

TABLE 8.4.—$\phi(0.01,\beta,n_1,n_2)$.—(Continued)

$$\beta = 0.975$$

| $n_2$ \ $n_1$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 15 | 20 | 24 | 30 | 40 | 60 | 120 | ∞ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 6.255 | 6.253 | 6.253 | 6.252 | 6.252 | 6.252 | 6.252 | 6.252 | 6.252 | 6.252 | 6.252 | 6.252 | 6.252 | 6.252 | 6.252 | 6.252 | 6.252 | 6.252 | 6.252 |
| 2 | 2.558 | 2.538 | 2.532 | 2.529 | 2.527 | 2.526 | 2.525 | 2.524 | 2.523 | 2.523 | 2.522 | 2.522 | 2.521 | 2.521 | 2.520 | 2.520 | 2.520 | 2.519 | 2.519 |
| 3 | 1.956 | 1.921 | 1.908 | 1.901 | 1.897 | 1.894 | 1.892 | 1.891 | 1.889 | 1.888 | 1.887 | 1.885 | 1.884 | 1.883 | 1.882 | 1.882 | 1.881 | 1.880 | 1.879 |
| 4 | 1.735 | 1.690 | 1.673 | 1.663 | 1.656 | 1.653 | 1.650 | 1.648 | 1.646 | 1.645 | 1.643 | 1.640 | 1.638 | 1.637 | 1.635 | 1.634 | 1.633 | 1.632 | 1.630 |
| 5 | 1.625 | 1.574 | 1.553 | 1.542 | 1.535 | 1.529 | 1.526 | 1.523 | 1.520 | 1.518 | 1.516 | 1.513 | 1.509 | 1.508 | 1.506 | 1.505 | 1.503 | 1.501 | 1.500 |
| 6 | 1.560 | 1.505 | 1.482 | 1.469 | 1.461 | 1.455 | 1.450 | 1.447 | 1.444 | 1.442 | 1.438 | 1.435 | 1.431 | 1.429 | 1.427 | 1.425 | 1.423 | 1.421 | 1.419 |
| 7 | 1.517 | 1.459 | 1.435 | 1.421 | 1.412 | 1.405 | 1.400 | 1.396 | 1.393 | 1.390 | 1.386 | 1.382 | 1.378 | 1.376 | 1.374 | 1.371 | 1.369 | 1.366 | 1.364 |
| 8 | 1.487 | 1.427 | 1.402 | 1.387 | 1.377 | 1.370 | 1.364 | 1.360 | 1.357 | 1.354 | 1.349 | 1.345 | 1.340 | 1.337 | 1.335 | 1.332 | 1.330 | 1.327 | 1.324 |
| 9 | 1.465 | 1.404 | 1.377 | 1.361 | 1.351 | 1.343 | 1.337 | 1.333 | 1.329 | 1.326* | 1.321 | 1.316 | 1.311 | 1.308 | 1.306 | 1.303 | 1.300 | 1.297 | 1.293 |
| 10 | 1.448 | 1.385 | 1.358 | 1.342 | 1.331 | 1.323 | 1.317 | 1.312 | 1.308 | 1.305 | 1.300 | 1.294 | 1.289 | 1.286 | 1.283 | 1.280 | 1.276 | 1.273 | 1.269 |
| 12 | 1.424 | 1.359 | 1.330 | 1.313 | 1.302 | 1.293 | 1.286 | 1.281 | 1.277 | 1.273 | 1.268 | 1.262 | 1.256 | 1.252 | 1.249 | 1.245 | 1.241 | 1.237 | 1.233 |
| 15 | 1.401 | 1.335 | 1.304 | 1.286 | 1.274 | 1.265 | 1.258 | 1.252 | 1.247 | 1.243 | 1.237 | 1.231 | 1.224 | 1.220 | 1.216 | 1.212 | 1.207 | 1.203 | 1.198 |
| 20 | 1.379 | 1.311 | 1.280 | 1.261 | 1.247 | 1.238 | 1.230 | 1.224 | 1.219 | 1.214 | 1.208 | 1.200 | 1.192 | 1.188 | 1.183 | 1.178 | 1.173 | 1.167 | 1.161 |
| 24 | 1.368 | 1.300 | 1.268 | 1.248 | 1.235 | 1.224 | 1.216 | 1.210 | 1.205 | 1.200 | 1.193 | 1.185 | 1.176 | 1.172 | 1.167 | 1.161 | 1.156 | 1.149 | 1.142 |
| 30 | 1.358 | 1.289 | 1.256 | 1.236 | 1.222 | 1.212 | 1.203 | 1.197 | 1.191 | 1.186 | 1.179 | 1.170 | 1.161 | 1.156 | 1.150 | 1.145 | 1.138 | 1.131 | 1.123 |
| 40 | 1.348 | 1.278 | 1.245 | 1.225 | 1.210 | 1.199 | 1.191 | 1.184 | 1.178 | 1.173 | 1.165 | 1.156 | 1.146 | 1.140 | 1.134 | 1.127 | 1.120 | 1.113 | 1.102 |
| 60 | 1.339 | 1.268 | 1.234 | 1.213 | 1.198 | 1.187 | 1.178 | 1.171 | 1.165 | 1.159 | 1.151 | 1.141 | 1.130 | 1.124 | 1.118 | 1.110 | 1.102 | 1.092 | 1.080 |
| 120 | 1.330 | 1.258 | 1.224 | 1.202 | 1.187 | 1.175 | 1.166 | 1.158 | 1.152 | 1.146 | 1.137 | 1.127 | 1.115 | 1.108 | 1.101 | 1.092 | 1.082 | 1.070 | 1.053 |
| ∞ | 1.321 | 1.248 | 1.214 | 1.191 | 1.176 | 1.164 | 1.154 | 1.146 | 1.139 | 1.133 | 1.123 | 1.112 | 1.099 | 1.092 | 1.083 | 1.073 | 1.061 | 1.044 | 1.000 |

For explanation see the panel of this table for $\beta = 0.005$.

TABLE 8.4   COMPARING TWO STANDARD DEVIATIONS   309

$$\beta = 0.995$$

| $n_2$ \ $n_1$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 12 | 15 | 20 | 24 | 30 | 40 | 60 | 120 | ∞ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.2500 | 0.2500 | 0.2500 | 0.2500 | 0.2500 | 0.2500 | 0.2500 | 0.2500 | 0.2500 | 0.2500 | 0.2500 | 0.2500 | 0.2500 | 0.2500 | 0.2500 | 0.2500 | 0.2500 | 0.2500 | 0.2500 |
| 2 | 0.4062 | 0.4975 | 0.4979 | 0.4981 | 0.4982 | 0.4983 | 0.4984 | 0.4984 | 0.4985 | 0.4985 | 0.4985 | 0.4986 | 0.4986 | 0.4986 | 0.4987 | 0.4987 | 0.4987 | 0.4987 | 0.4987 |
| 3 | 0.6141 | 0.6188 | 0.6206 | 0.6215 | 0.6221 | 0.6225 | 0.6228 | 0.6230 | 0.6231 | 0.6233 | 0.6235 | 0.6237 | 0.6239 | 0.6240 | 0.6241 | 0.6243 | 0.6244 | 0.6245 | 0.6246 |
| 4 | 0.6765 | 0.6848 | 0.6882 | 0.6900 | 0.6912 | 0.6920 | 0.6926 | 0.6931 | 0.6935 | 0.6938 | 0.6943 | 0.6947 | 0.6952 | 0.6954 | 0.6957 | 0.6959 | 0.6961 | 0.6964 | 0.6967 |
| 5 | 0.7135 | 0.7248 | 0.7296 | 0.7323 | 0.7341 | 0.7353 | 0.7363 | 0.7370 | 0.7376 | 0.7381 | 0.7388 | 0.7396 | 0.7403 | 0.7407 | 0.7411 | 0.7415 | 0.7420 | 0.7424 | 0.7428 |
| 6 | 0.7376 | 0.7512 | 0.7571 | 0.7606 | 0.7629 | 0.7646 | 0.7658 | 0.7668 | 0.7676 | 0.7682 | 0.7692 | 0.7702 | 0.7713 | 0.7719 | 0.7724 | 0.7730 | 0.7736 | 0.7742 | 0.7748 |
| 7 | 0.7542 | 0.7696 | 0.7766 | 0.7806 | 0.7835 | 0.7855 | 0.7870 | 0.7882 | 0.7892 | 0.7900 | 0.7912 | 0.7925 | 0.7938 | 0.7945 | 0.7953 | 0.7960 | 0.7968 | 0.7976 | 0.7984 |
| 8 | 0.7665 | 0.7833 | 0.7910 | 0.7957 | 0.7988 | 0.8011 | 0.8029 | 0.8043 | 0.8054 | 0.8063 | 0.8078 | 0.8093 | 0.8110 | 0.8118 | 0.8127 | 0.8136 | 0.8145 | 0.8155 | 0.8165 |
| 9 | 0.7757 | 0.7937 | 0.8021 | 0.8072 | 0.8107 | 0.8133 | 0.8152 | 0.8168 | 0.8181 | 0.8191 | 0.8208 | 0.8225 | 0.8244 | 0.8254 | 0.8264 | 0.8275 | 0.8286 | 0.8298 | 0.8309 |
| 10 | 0.7831 | 0.8019 | 0.8109 | 0.8164 | 0.8201 | 0.8229 | 0.8251 | 0.8268 | 0.8282 | 0.8294 | 0.8312 | 0.8332 | 0.8353 | 0.8364 | 0.8376 | 0.8388 | 0.8400 | 0.8414 | 0.8427 |
| 12 | 0.7938 | 0.8140 | 0.8238 | 0.8299 | 0.8342 | 0.8373 | 0.8398 | 0.8418 | 0.8434 | 0.8448 | 0.8469 | 0.8493 | 0.8518 | 0.8531 | 0.8545 | 0.8560 | 0.8575 | 0.8591 | 0.8609 |
| 15 | 0.8041 | 0.8257 | 0.8365 | 0.8432 | 0.8480 | 0.8516 | 0.8544 | 0.8567 | 0.8586 | 0.8601 | 0.8627 | 0.8654 | 0.8684 | 0.8701 | 0.8719 | 0.8736 | 0.8755 | 0.8776 | 0.8798 |
| 20 | 0.8142 | 0.8372 | 0.8488 | 0.8563 | 0.8616 | 0.8657 | 0.8689 | 0.8715 | 0.8737 | 0.8755 | 0.8785 | 0.8818 | 0.8854 | 0.8875 | 0.8896 | 0.8918 | 0.8943 | 0.8970 | 0.8999 |
| 24 | 0.8190 | 0.8428 | 0.8549 | 0.8627 | 0.8683 | 0.8726 | 0.8761 | 0.8789 | 0.8812 | 0.8832 | 0.8865 | 0.8900 | 0.8941 | 0.8963 | 0.8987 | 0.9012 | 0.9041 | 0.9072 | 0.9107 |
| 30 | 0.8238 | 0.8483 | 0.8608 | 0.8691 | 0.8750 | 0.8795 | 0.8832 | 0.8862 | 0.8887 | 0.8909 | 0.8944 | 0.8984 | 0.9028 | 0.9053 | 0.9080 | 0.9109 | 0.9142 | 0.9178 | 0.9220 |
| 40 | 0.8285 | 0.8536 | 0.8667 | 0.8753 | 0.8815 | 0.8864 | 0.8903 | 0.8935 | 0.8962 | 0.8985 | 0.9024 | 0.9067 | 0.9117 | 0.9145 | 0.9175 | 0.9209 | 0.9247 | 0.9291 | 0.9342 |
| 60 | 0.8331 | 0.8589 | 0.8725 | 0.8814 | 0.8880 | 0.8931 | 0.8973 | 0.9007 | 0.9037 | 0.9062 | 0.9104 | 0.9151 | 0.9207 | 0.9238 | 0.9274 | 0.9313 | 0.9358 | 0.9412 | 0.9479 |
| 120 | 0.8376 | 0.8641 | 0.8781 | 0.8875 | 0.8944 | 0.8998 | 0.9043 | 0.9079 | 0.9111 | 0.9138 | 0.9184 | 0.9236 | 0.9298 | 0.9335 | 0.9375 | 0.9422 | 0.9478 | 0.9548 | 0.9646 |
| ∞ | 0.8431 | 0.8692 | 0.8837 | 0.8934 | 0.9007 | 0.9064 | 0.9111 | 0.9151 | 0.9185 | 0.9214 | 0.9264 | 0.9322 | 0.9392 | 0.9434 | 0.9482 | 0.9540 | 0.9612 | 0.9713 | 1.000 |

For explanation see the panel of this table for $\beta = 0.005$.

manner, and the material produced exhibits an inherent random variability. This random variation may have assignable components. For instance, the thickness of washers punched from a single sheet may be regarded as varying at random about the mean thickness of washers from that sheet, because of the inherent nonuniformity of the sheet; and washers produced during a given shift may have still greater variation about the mean thickness for the shift, because of an additional component of variation assignable to the random variation of the mean thicknesses of the respective sheets about the over-all mean thickness of the sheets on that shift, and so forth.

(2) There are factors that introduce *systematic* variation. For instance, in a production unit composed of several similar machines producing the same product simultaneously, there are generally systematic differences among the machines, which superpose a pattern of variation on the output of that production unit.

In using analysis of variance merely to determine whether the over-all variation has a component assignable to a particular factor, the computational procedure and the mechanics of the statistical test are the same in the case of a factor that may contribute a component of random variation as in the case of a factor that may contribute a component of systematic variation. On the other hand, in planning such investigations and in the estimation of the magnitude(s) of the component(s), the distinction has to be taken into account, since the situations between which discrimination is required differ in the two cases, leading to differences in the computation of probabilities of errors of the second kind. *We consider only cases where random variation alone is present.*

The question arises as to how to decide in a given case whether a component of variation is systematic or random. The answer depends upon the extent to which the experimental procedure samples the material at random. If a sampling operation is such that each item of the population has an equal chance of being included in the sample, and separate drawings are independent except for any dependence resulting from the number of items in the population being limited, then the sampling is random and the sample may be termed a random sample.* When an experimenter selects two or more treatments for comparison in some respect, he rarely, if ever, selects them at random from a population of possible treatments; he selects those he believes most promising. Accordingly, components of variation assignable to differences between treatments are systematic. On the other hand, when an experimenter

---

* It is important to note that it is the operation of drawing the sample that is random and not the sample itself, since any sample whatsoever can be obtained by either a random or a nonrandom operation.

selects a sample of items from a production lot (or a sample of animals from a herd) for a study of the effects of the treatments, he can ensure that they are a random sample from the lot, or herd, by introducing randomization into the sampling procedure, for example, by drawing cards from a hat after thorough mixing, or by using a table of random numbers. He may consider such a sample to be a random sample from the production process (or the animal species) only by making the assumption that the lot (herd) itself is a random sample from the process (species). When a particular experiment is being planned, or the results of a completed experiment are being interpreted, the following two parallel sets of questions will serve to focus attention on the issue of systematic versus random variation:

(1) Are the conclusions to be confined to the things actually studied (the sample of items, or animals), or to the immediate sources of these things (the actual machines, or lots, or herds from which the things studied were obtained); or are they to be expanded to apply to the underlying populations (real or hypothetical population of machines, or the production process, or the species)?

(2) In complete repetitions of the experiment, would the same things be studied again (the same items, animals, or treatments); would new samples be drawn from the identical sources (the same machines, the same lots, or the same herds); or would new samples be drawn from the underlying populations (samples of items from different machines, or animals from different herds of the same breed)?

### 5.2. A Typical Analysis of Variance

This section gives a typical analysis-of-variance table for cases in which the random deviations involved are normally distributed. The conditions are specified under which the test of significance customarily associated with this table can be regarded as a test of the hypothesis that two variances $\sigma_1^2$ and $\sigma_2^2$ are equal, relative to the alternative that $\sigma_1^2 = \sigma_2^2 + K\sigma_3^2$, where $K$ is a known constant and $\sigma_3^2$ is an unknown variance. Under these conditions the results given in Sec. 4.1 can be used in the planning and interpretation of experiments leading to analysis of variance tests of significance.

Let $x_{ij}$ denote the $j$th observation in the $i$th group ($i = 1, 2, \cdots , k; j = 1, 2, \cdots , n_i$), let $x_i.$ denote the average of the $n_i$ observations in the $i$th group ($i = 1, 2, \cdots , k$), and let $x..$ denote the average of all $N = \sum_{i=1}^{k} n_i$ observations; then the analysis-of-variance table takes the following form:

| Variation | Degrees of freedom | Sum of squares | Mean square | Expectation of mean square |
|---|---|---|---|---|
| Between group averages.... | $k - 1$ | $S_b^2 = \displaystyle\sum_{i=1}^{k} n_i (x_{i\cdot} - x_{\cdot\cdot})^2$ | $S_b^2/(k - 1) = s_b^2$ | $\sigma^2 + \bar{n}\sigma_g^2$ |
| Within groups. | $N - k$ | $S_w^2 = \displaystyle\sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - x_{i\cdot})^2$ | $S_w^2/(N - k) = s_w^2$ | $\sigma^2$ |

$$\bar{n} = \left(N^2 - \sum_{i=1}^{k} n_i^2\right)/[N(k - 1)]; \text{ when } n_1 = n_2 = \cdots n_k = n, \bar{n} = n.$$

If the $x_{ij}$ are independently and normally distributed about the respective true means $m_i$ of the groups with common (unknown) variance $\sigma^2$, and if the groups are a random selection from a population of groups such that in repeated selection the $m_i$ would be normally distributed about an over-all population mean of $m$ with variance $\sigma_g^2$, the $x_{ij}$ will be normally distributed about $m$ with variance

$$\sigma_x^2 = \sigma^2 + \sigma_g^2 \tag{29}$$

likewise

$$\frac{(k - 1)s_b^2}{\sigma^2 + \bar{n}\sigma_g^2} \tag{30}$$

will be distributed as $\chi^2$ for $(k - 1)$ degrees of freedom; and finally

$$\frac{(N - k)s_w^2}{\sigma^2} \tag{31}$$

will be distributed independently as $\chi^2$ for $(N - k)$ degrees of freedom.

A comparison of (30) and (31) with the assumptions made with regard to $s_1^2$ and $s_2^2$ in Sec. 4.1 reveals that if we set

$$\sigma_1^2 = \sigma^2 + \bar{n}\sigma_g^2 \quad \text{and} \quad \sigma_2^2 = \sigma^2 \tag{32}$$

$$n_1 = (k - 1) \quad \text{and} \quad n_2 = N - k \tag{33}$$

$s_b^2$ becomes identified with $s_1^2$ and $s_w^2$ with $s_2^2$, and the rule of procedure becomes

(a) Accept $\sigma_g^2 = 0$ (there exists no between-group component of random variation) whenever

$$\frac{s_b^2}{s_w^2} \leq F_\alpha(k - 1, N - k) \tag{34}$$

(b)  Accept $\sigma_g^2 > 0$ (there exists a between-group component of random variation) whenever

$$\frac{s_b^2}{s_w^2} > F_\alpha(k - 1, N - k) \tag{35}$$

Since $\sigma_g^2$ cannot be negative, the probability of an error of the first kind (concluding that $\sigma_g^2 > 0$ when actually $\sigma_g^2 = 0$) is exactly $\alpha$, and the probability of an error of the second kind (concluding that $\sigma_g^2 = 0$ when actually $\sigma_g^2 > 0$) is $\beta$ when

$$1 + \bar{n}\left(\frac{\sigma_g^2}{\sigma^2}\right) = \phi(\alpha, \beta, k - 1, N - k) \tag{36}$$

If there are $n$ observations in each group, $\bar{n} = n$ and

$$N - k = k(n - 1);$$

so*

$$\frac{\sigma_g^2}{\sigma^2} = \frac{\phi[\alpha, \beta, k - 1, k(n - 1)] - 1}{n} \tag{37}$$

## 5.3. Illustrations

The following examples illustrate types of questions to which answers may be obtained with the aid of (37):

*Example* VII.  It is suspected that differences between lots contribute to the random variation of a particular characteristic of a certain product.  To investigate this hypothesis, $n$ items are to be selected at random and measured, from each of $k$ lots chosen at random from the lots currently accessible.  It is agreed that the test of significance defined by (34) and (35) is to be conducted at the 0.05 level of significance, and it is desired to have a probability not greater than 0.50 of falsely concluding $\sigma_g^2 = 0$ if $\sigma_g^2/\sigma^2 = 0.10$.  What combination of values of $k$ and $n$ will accomplish this with a minimum total number of measurements?

Baines has prepared the following table for answering such questions when $\alpha = 0.05$ and $\beta = 0.50$:

| $k$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 14 | 17 | 21 | 27 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | 75 | 34 | 24 | 19 | 16 | 14 | 13 | 12 | 11 | 9 | 8 | 7 | 6 |
| $kn$ | 150 | 102 | 96 | 95 | 96 | 98 | 104 | 108 | 110 | 126 | 136 | 147 | 162 |
| $\sigma_g^2/\sigma^2$ | 0.101 | 0.101 | 0.100 | 0.100 | 0.103 | 0.104 | 0.101 | 0.100 | 0.102 | 0.099 | 0.099 | 0.100 | 0.102 |

It is seen that the best combination is 19 items from each of 5 lots, requiring a total of 95 measurements.  However, if it is more con-

---

* BAINES, *Economical Design*, tabulates (37) for $\alpha = 0.05$, $\beta = 0.50$, $k = 2$ to 31 by steps of 1, and $n = 1$ to 31 by steps of 1, 31 to 45 by steps of 2, 45 to 93 by steps of 4, 101, and 121.

venient to use 4 lots than 5 lots, 24 items from each of 4 lots will be equally effective and will require only one additional measurement.

*Example* V.III. If it were agreed initially in planning a test of the type discussed in Example VII that a total of 96 measurements would be employed, then the problem would arise of determining the combination of values of $k$ and $n$, subject to the restriction $kn = 96$, that minimizes the value of $\sigma_0^2/\sigma^2$ for a given $\beta$.

Baines gives the figures tabulated below, which apply when $\alpha = 0.05$ and $\beta = 0.50$:

| $k$ | 2 | 3 | 4 | 6 | 8 | 12 | 16 | 24 |
|---|---|---|---|---|---|---|---|---|
| $n$ | 48 | 32 | 24 | 16 | 12 | 8 | 6 | 4 |
| $\sigma_0^2/\sigma^2$ | 0.158 | 0.107 | 0.100 | 0.103 | 0.110 | 0.126 | 0.143 | 0.178 |

It is clear that the combination $k = 4$, $n = 24$ provides the most stringent test having $\alpha = 0.05$ and $kn = 96$, in the sense that it yields the minimum indifference point.

In order to answer readily questions of the type illustrated by Examples VII and VIII when $\alpha$ and $\beta$ differ from 0.05 and 0.50, respectively, tables of (37) should be prepared using $k$ and $n$ as arguments. Indeed, to facilitate the planning of experiments leading to analysis-of-variance tests corresponding to two-way classifications ("rows" and "columns"), to Latin square designs, and so forth, sets of such tables will have to be prepared for each type of design, since the number of degrees of freedom corresponding to the estimate of error variance $s_2^2$ depends upon the design of the experiment [for a two-way classification, for example, $n_2 = (r - 1)(c - 1)$], as does also the number of degrees of freedom corresponding to the estimate of variance $s_1^2$ to be tested. Until such tables are available, these questions can be answered, with considerable effort expended on interpolation, by means of Tables 8.3 and 8.4. Furthermore, Tables 8.3 and 8.4 will often suffice to determine a few points on the operating-characteristic curve of a test of significance corresponding to an experimental design already agreed upon.

## 6. TECHNICAL NOTES

### 6.1. Method of Construction and Accuracy of Tables 8.1 and 8.2

As explained in Sec. 2, the standard procedure for determining whether a new product or process is more variable than a standard product or process is to compute the statistic $s_1^2/\sigma_0^2$ and

(*a*) Accept the conclusion $\sigma_1^2/\sigma_0^2 \leq 1$ whenever

$$\frac{s_1^2}{\sigma_1^2} \leq \frac{\chi_\alpha^2(n_1)}{n_1} \tag{38}$$

(b) Accept the conclusion $\sigma_1^2/\sigma_0^2 > 1$ whenever

$$\frac{s_1^2}{\sigma_0^2} > \frac{\chi_\alpha^2(n_1)}{n_1} \tag{39}$$

where $\chi_\alpha^2(n_1)$ is the $\alpha$ level of significance of $\chi^2$ for $n_1$ degrees of freedom. The level of significance $\alpha$ represents the probability of accepting the conclusion $\sigma_1^2/\sigma_0^2 > 1$ when in fact $\sigma_1^2/\sigma_0^2 = 1$ (error of the first kind). Using this level of significance, the probability of accepting the conclusion $\sigma_1^2/\sigma_0^2 \leq 1$ when in fact $\sigma_1^2/\sigma_0^2 > 1$ (error of the second kind) depends on the ratio $\rho = \sigma_1^2/\sigma_0^2$ and is exactly $\beta$ if $\rho$ has the value

$$\rho(\alpha,\beta,n_1) = \frac{\chi_\alpha^2(n_1)}{\chi_{1-\beta}^2(n_1)} \tag{40}$$

The proof is elementary. Since the conclusion $\sigma_1^2/\sigma_0^2 \leq 1$ will be reached whenever (38) occurs, the probability of this conclusion, given $\sigma_1^2 = \rho\sigma_0^2$, is

$$P\left\{\frac{s_1^2}{\sigma_0^2} \leq \frac{\chi_\alpha^2(n_1)}{n_1}\middle|\sigma_1^2 = \rho\sigma_0^2\right\} = P\left\{\frac{n_1 s_1^2}{\sigma_1^2} \leq \frac{\sigma_0^2}{\sigma_1^2}\chi_\alpha^2(n_1)\middle|\sigma_1^2 = \rho\sigma_0^2\right\} \tag{41}$$

$$= P\left\{\chi^2(n_1) \leq \frac{\chi_\alpha^2(n_1)}{\rho}\right\}$$

$$= 1 - P\left\{\chi^2(n_1) > \frac{\chi_\alpha^2(n_1)}{\rho}\right\}$$

From (40),

$$\frac{\chi_\alpha^2(n_1)}{\rho} = \chi_{1-\beta}^2(n_1) \tag{42}$$

so the last form of (41) becomes

$$1 - (1 - \beta) = \beta \tag{43}$$

which completes the proof.*

Thompson's table of $\chi_\alpha^2(n_1)$ was used in constructing Tables 8.1 and 8.2. Direct substitution of these values in formula (40) yields values of $\rho(\alpha,\beta,n)$ corresponding to $\alpha = 0.05$ and 0.01 and to $\beta = 0.005$, 0.01, 0.025, 0.05, 0.10, 0.25, 0.50, 0.750, 0.900, 0.950, 0.975, 0.990, and 0.995 for $n = 1$ to 30 by steps of 1 and 30 to 100 by steps of 10. The values of $\rho(0.05,\beta,n)$, rounded to 5 figures or 4 decimals, are given in

* Values of $\chi_\alpha^2(n_1)$ corresponding to $\alpha = 0.01$, 0.02, 0.05, 0.10, 0.20, 0.30, 0.50, 0.70, 0.80, 0.90, 0.95, 0.98, and 0.99 are given to 3 decimals for $n_1 = 1$ to 30 by steps of 1 in Table III of Fisher, *Statistical Methods* and in Table IV of Fisher and Yates, *Statistical Tables*, the latter of which also contains the values for $\alpha = 0.001$. Values of $\chi_\alpha^2(n_1)$ corresponding to $\alpha = 0.005$, 0.010, 0.025, 0.050, 0.100, 0.250, 0.500, 0.750, 0.900, 0.950, 0.975, 0.990, and 0.995 are given to 6 significant figures for $n_1 = 1$ to 30 by steps of 1 and from 30 to 100 by steps of 10 by Thompson, "$\chi^2$ distribution."

Table 8.1.  The corresponding values of $\rho(0.01,\beta,n)$ are given in Table 8.2.  The values of $\rho(\alpha,\beta,n)$ for $n = 120$ were computed from values of $\chi_\alpha^2(120)$ obtained from the percentage points of the $F$ distribution tabled by Merrington and Thompson, by means of the relation

$$\chi_\alpha^2(n) = nF_\alpha(n, \infty) \tag{44}$$

Two-point interpolation in Tables 8.1 and 8.2 in terms of the argument $1/\sqrt{n}$ will yield about 2-decimal accuracy in most parts of these tables.*

### 6.2.  Method of Construction and Accuracy of Tables 8.3 and 8.4

The procedure described in Sec. 4.1 for determining whether a particular one of two new products or processes is more variable than the other is as follows: Compute the statistic $s_1^2/s_2^2$, where $s_1^2$ and $s_2^2$ are sample estimates of the two population variances such that in repeated sampling $n_1 s_1^2/\sigma_1^2$ and $n_2 s_2^2/\sigma_2^2$ will be distributed independently as $\chi^2$ for $n_1$ and $n_2$ degrees of freedom, respectively.  Base the choice of product or process on the rule

(a)  Accept product I whenever

$$\frac{s_1^2}{s_2^2} \leq F_\alpha(n_1,n_2) \tag{45}$$

(b)  Accept product II whenever

$$\frac{s_1^2}{s_2^2} > F_\alpha(n_1,n_2) \tag{46}$$

where $F_\alpha(n_1,n_2)$ is the $\alpha$ level of significance of the variance ratio $(F)$ distribution.  This rule is so formulated that the probability of accepting product II when in fact $\sigma_1^2/\sigma_2^2 = 1$ (error of the first kind) equals the level of significance $\alpha$.  The probability of accepting product I when in fact $\sigma_1^2 > \sigma_2^2$ (error of the second kind) depends on the ratio $\phi = \sigma_1^2/\sigma_2^2$ and is exactly $\beta$ when

$$\phi(\alpha,\beta,n_1,n_2) = \frac{F_\alpha(n_1,n_2)}{F_{1-\beta}(n_1,n_2)} \tag{47}$$
$$= F_\alpha(n_1,n_2) \cdot F_\beta(n_2,n_1)$$

The proof is similar to the proof of (40).  Product I will be chosen whenever (45) occurs, and the probability of this when $\sigma_1^2/\sigma_2^2 = \phi$ is

---

* Interpolation with respect to $\alpha$ and $\beta$ should be carried out in terms of $K_\alpha$ and $K_\beta$, respectively, where $K_\epsilon$ is the unit normal deviate exceeded with probability $\epsilon$ (tabled in Chap. 1, Table 1.1).  For an explanation and examples of this type of interpolation, see Simaika, *Interpolation*.

$$P\left\{\frac{s_1^2}{s_2^2} \leq F_\alpha(n_1,n_2)\middle|\frac{\sigma_1^2}{\sigma_2^2} = \phi\right\} = P\left\{\frac{\frac{s_1^2}{\sigma_1^2}}{\frac{s_2^2}{\sigma_2^2}} \leq \frac{F_\alpha(n_1,n_2)}{\phi}\middle|\frac{\sigma_1^2}{\sigma_2^2} = \phi\right\} \qquad (K)$$

$$= P\left\{F(n_1,n_2) \leq \frac{F_\alpha(n_1,n_2)}{\phi}\right\}$$

$$= 1 - P\left\{F(n_1,n_2) > \frac{F_\alpha(n_1,n_2)}{\phi}\right\}$$

$$= 1 - (1 - \beta) = \beta$$

when*

$$\frac{F_\alpha(n_1,n_2)}{\phi} = F_{1-\beta}(n_1,n_2) \qquad (49)$$

Tables 8.3 and 8.4 at the end of this chapter were computed from the Merrington-Thompson table of $F$ by direct substitution in formula (47). These tables give the values of $\phi(\alpha,\beta,n_1,n_2)$ for $\alpha = 0.050$ and $\alpha = 0.010$, respectively, corresponding to $\beta = 0.005, 0.010, 0.025, 0.050, 0.100, 0.250, 0.500, 0.750, 0.900, 0.950, 0.975, 0.990, 0.995$, for $n_1 = 1$ to $10$ by steps of 1, and 12, 15, 20, 24, 30, 40, 60, 120, $\infty$, and $n_2 = 1$ to 10 by steps of 1, and 12, 15, 20, 24, 30, 40, 60, 120, $\infty$. Interpolation for values of $\phi(\alpha,\beta,n_1,n_2)$ corresponding to values of $n_1$ and $n_2$ between those given in these tables should be done harmonically, that is, in terms of $1/n_1$ and $1/n_2$. In this connection it should be noted that the sequence $\infty$, 120, 60, 40, 30, 24, 20 corresponds to $120/m$ for $m = 0, 1, 2, 3, 4, 5$, and 6, respectively; and that the sequence $\infty$, 60, 30, 20, 15, 12, 10 corresponds to $60/r$ for $r = 0, 1, 2, 3, 4, 5$, and 6, respectively. Accordingly, interpolation is most conveniently carried out in terms of the arguments $m$ and $r$. Two-point harmonic interpolation will generally give $\phi$ to sufficient accuracy for most practical purposes. For example, by 2-point harmonic interpolation

$$\phi(0.05,0.05,30,24) \doteq \frac{\phi(0.05,0.05,30,20) + \phi(0.05,0.05,30,30)}{2} = 3.664$$

$$(50)$$

The actual value is 3.660; customary 2-point linear interpolation gives

---

* The most accurate table of $F_\alpha(n_1,n_2)$ is that of Merrington and Thompson, which gives the values to 5 significant figures corresponding to $\alpha = 0.005, 0.010, 0.025, 0.050, 0.100, 0.250$, and 0.500 for $n_1 = 1$ to 10 by steps of 1, and 12, 15, 20, 24, 30, 40, 60, 120, $\infty$, and $n_2 = 1$ to 30 by steps of 1, and 40, 60, 120, $\infty$. Values of $F_\alpha(n_1,n_2)$, referred to as $e^{2z}$, corresponding to $\alpha = 0.001, 0.01, 0.05, 0.20$, are given (mostly to 2 decimals) in Table V of Fisher and Yates, *Statistical Tables* for $n_1 = 1$ to 6 by steps of 1, and 8, 12, 24, $\infty$, and $n_2 = 1$ to 30 by steps of 1, and 40, 60, 120, $\infty$.

3.719. Similarly, by 2-point harmonic interpolation

$$\phi(0.05,0.05,30,120) \doteq \frac{\phi(0.05,0.05,30,60) + \phi(0.05,0.05,30,\infty)}{2} = 2.618$$

(51)

The actual value is 2.617. Use may be made of the fact that

$$\phi(\alpha,\beta,n_1,\infty) = \rho(\alpha,\beta,n_1)$$

(52)

to obtain values of the left side of this equality for values of $n_1$ included in Tables 8.1 and 8.2 but not in Tables 8.3 and 8.4.

### 7. REFERENCES*

BAINES, A. H. J.: *On Economical Design of Statistical Experiments*, (British) Ministry of Supply, Advisory Service on Statistical Method and Quality Control, Technical Report, Series *R*, No. Q.C./R./15, July 15, 1944.

DANIELS, H. E.: "The Estimation of Components of Variance," *Supplement to the Journal of the Royal Statistical Society*, Vol. 6 (1939), pp. 186–197.

FISHER, R. A.: *Statistical Methods for Research Workers*, 5th and later ed., Oliver and Boyd, Ltd., Edinburgh and London, 1934–1944.

———, and YATES, FRANK: *Statistical Tables for Biological, Agricultural and Medical Research*, 2d ed., Oliver and Boyd, Ltd., London and Edinburgh, 1943, 98 + viii pp.

MERRINGTON, MAXINE, and THOMPSON, CATHERINE M.: "Tables of Percentage Points of the Inverted Beta (*F*) Distribution," *Biometrika*, Vol. 33 (1943), pp. 73–88.

NEYMAN, J., and PEARSON, E. S.: "On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference," *Biometrika*, Vol. 20A (1928), pp. 175–240, 263–294.

SIMAIKA, J. B.: "Interpolation for Fresh Probability Levels between the Standard Table Levels of a Function," *Biometrika*, Vol. 32 (1941), pp. 263–276.

THOMPSON, CATHERINE M.: "Table of Percentage Points of the $\chi^2$ Distribution," *Biometrika*, Vol. 32 (1941), pp. 187–191.

* While this volume was in press there appeared a paper by C. D. Ferris, F. E. Grubbs, and C. L. Weaver, "Operating Characteristics for the Common Statistical Tests of Significance," *Annals of Mathematical Statistics*, Vol. 17 (1946), pp. 178–197. Fig. 1, p. 181, of the paper gives operating-characteristic curves similar to those of Fig. 8.1 of this chapter, for $n_1 = 2$ to 10, 15, 20 to 50 by steps of 10, 75, and 100. Curves similar to those of Fig. 8.3 are given in Fig. 3, p. 185, for $n_1 = n_2 = 2$ to 10, 15, 20 to 50 by steps of 10, 75, and 100, in Fig. 4, p. 186, for certain cases in which $3n_1 = 2n_2$ or $2n_1 = n_2$, and in Fig. 5, p. 187, for certain cases in which $2n_1 = 3n_2$ or $n_1 = 2n_2$.

# CHAPTER 9

## UTILIZATION OF LIMITED EXPERIMENTAL FACILITIES WHEN THE COST OF EACH MEASUREMENT DEPENDS ON ITS MAGNITUDE

This chapter was prepared by
Milton Friedman, University of Chicago

# CHAPTER 9

## UTILIZATION OF LIMITED EXPERIMENTAL FACILITIES WHEN THE COST OF EACH MEASUREMENT DEPENDS ON ITS MAGNITUDE

### 1. TYPE OF PROBLEM

A primary objective in the statistical design of experiments is to maximize the amount of pertinent information obtained from a given amount of time, equipment, personnel, and supplies, or, conversely, to minimize the cost of obtaining a given amount of information. In many experiments the cost of an observation can be regarded as a constant; such is the case, for example, in the problem of Chap. 11. Sometimes the cost of an observation is determined by the independent variables, as, for example, when an independent variable is the amount of some expensive substance or type of energy. This chapter is concerned with a third situation, in which the cost of an observation depends on the value of the dependent variable; a simple case occurs when the dependent variable is determined by counting, so that the time and labor involved in a measurement are proportional to its magnitude. Insofar as the dependent and independent variables are correlated, the latter two situations are essentially the same.

An interesting case in which the cost of an observation depends on its magnitude, one which illustrates the interplay of statistical principles and substantive considerations in applied statistics, occurs in metallurgical experiments to determine the relation between time to rupture and stress for a given alloy. In investigating this relation, each observation is obtained by subjecting a test specimen to a predetermined constant stress and recording the amount of time that elapses before the specimen breaks. The number of observations that can be obtained in a given period of time from a single piece of equipment thus depends on how long it takes each specimen to rupture, which in turn depends largely on the stress to which the test specimen is subjected.

If the logarithm of time to rupture is a linear function of the logarithm of stress (and there is strong evidence that this is approximately so for a large number of alloys over the range of important stresses), tests at any two stresses suffice to estimate the line relating the two variables, and from this line can be predicted the time to rupture at any other stress. The error in such a prediction depends on the number of tests from which the line is estimated and on the stresses at which they are

made.   But the number of tests made itself depends on the stresses at which they are made, since the lower the stress, the longer it takes for a specimen to break, and hence the fewer the tests that can be made in a limited period with given equipment.   This factor, taken by itself, would argue for using as high stresses as possible, in order to have many tests. However, the higher the stresses, the farther it is necessary to extrapolate the line to predict time to rupture at the stresses which are practically important.   Further, while the percentage error in test results seems reasonably independent of the stress used for a fairly wide range of stresses, it will doubtless increase if the stress is raised sufficiently. Thus there is a real problem about the best stresses at which to test; there are opposing considerations that have to be evaluated quantitatively to get a definite answer.

## 2. SPECIFIC PROBLEM

It will facilitate the solution of the problem to state it more formally, as follows:

*Given:* (1)  The relation between time to rupture and stress is

$$t = \alpha + \beta s \tag{1}$$

where $t$ is the logarithm of time to rupture, $s$ is the logarithm of the stress, and $\alpha$ and $\beta$ are constants describing the "true" (but unknown) relationship.

(2)  The number of specimens that can be tested is limited solely by the fact that there are a limited number of machines available and a limited amount of time.

(3)  The percentage error in the determination of time to rupture does not depend on stress; that is, if a great many determinations were made at each of a number of stresses, the percentage variation in the determinations would be the same at all stresses.   This assumption will undoubtedly break down at sufficiently high stresses, but it seems to be confirmed by the available evidence for stresses ordinarily used.

*Required:* (1)  The stresses at which to make the tests, and (2) the best apportionment of the tests among these stresses, in order to estimate as accurately as possible (a) the slope $\beta$ of the line relating $t$ to $s$, and (b) the time to rupture at some predetermined stress, or range of stresses, of practical importance, ordinarily considerably below the stresses at which the tests are conducted.

## 3. SOLUTION

The solution of this problem is as follows:

(1)  It is always better to use two testing stresses than to use more than two.

(2) For purpose (*b*) it might sometimes be best to use only one stress $S_0$ of practical importance. However, since this would give no information about the time to rupture at any other stress, it would ignore purpose (*a*) and is therefore excluded.

(3) Let $S_1$ and $S_2$ be the two best testing stresses $(S_2 > S_1)$. Then the best ratio of the two stresses, say, $R = S_1/S_2$, is given by solving the following equation for $R$:

$$\frac{\beta}{2} \log_e R = 1 + R^{-\frac{\beta}{2}} \tag{2}$$

This is the best ratio both for estimating the slope and for estimating the time to rupture at $S_0$, provided $S_0$ is less than $S_1$, the case of primary interest.* The best ratio of the two testing stresses is therefore independent of $S_0$ and depends only on $\beta$. The value of $\beta$ is, of course, to be determined from the experiment; but some advance estimate is needed because the relative times required for tests at two different stresses depend on $\beta$. However, the advance estimate need not be very accurate, since even fairly large errors in the estimate of $\beta$ will not lead to any great loss in efficiency.

(4) Under the given assumptions, the absolute values of $S_1$ and $S_2$ cannot be determined. So long as assumption (3) of Sec. 2 is satisfied, namely, that the percentage error is independent of stress, the higher the testing stresses the better. In practice, therefore, $S_2$ should be chosen as the highest testing stress for which the percentage error can reasonably be expected to be no larger than at lower stresses. $S_1$ can then be determined from the $R$ given by equation (2).

(5) Let $n_1$ be the number of tests to be made at $S_1$ and $n_2$ the number to be made at $S_2$. Let $r = n_1/n_2$. The ratio $r$ is all that needs to be determined, since the total number of tests, $n_1 + n_2$, is determined by $S_1$ and $S_2$ and by the time and equipment available. The best value of $r$ depends on the objective. If the objective is to estimate $\beta$ as accurately as possible, the best value of $r$ is given by

$$r = R^{-\frac{\beta}{2}} \tag{3}$$

If the objective is to estimate the time to rupture at $S_0$ as accurately as possible, the best value of $r$ is given by

$$r = R^{-\frac{\beta}{2}} \frac{\log \frac{S_0}{S_2}}{\log \frac{S_0}{S_1}} \tag{4}$$

---

* See Sec. 4.3, especially the last paragraph, for a discussion of the case where $S_1 \leq S_0$.

The best value of $r$ in this case depends on $S_0$. This means that it is not possible to design the experiment so as to get the best estimate of the time to rupture at each of a range of $S_0$'s. It is necessary to compromise among the objectives.

(6) The following numerical calculation illustrates the solution. For one particular series of cast alloys tested at a particular elevated temperature, $\beta$ is found to be approximately $-8$. If $S_2$ is taken as 30,000 pounds per square inch (p.s.i.), the best value of $S_1$ turns out to be 22,000 p.s.i. The best apportionment of the tests between the two stresses is as follows:

| Objective is to estimate as accurately as possible | Number of tests to be made at 22,000 p.s.i. for each test at 30,000 p.s.i. | Percentage gain in precision over using 0.50 tests at 25,000 p.s.i. for every test at 30,000 p.s.i. |
|---|---|---|
| (a)  Slope...................... | 0.26 | 39 |
| (b)  Time to rupture at stress of | | |
| 5,000 p.s.i............... | 0.35 | 42 |
| 10,000 p.s.i............... | 0.41 | 45 |
| 15,000 p.s.i............... | 0.53 | 47 |
| 20,000 p.s.i............... | 1.25 | 55 |

The last column gives the gain in precision, which is equivalent to the savings in testing for a given degree of precision, over an alternative combination of stresses and relative number of observations, which differs only moderately from the optimum combination; it indicates the value of careful planning of the stresses at which to test and of the relative number of tests at different stresses.

### 4. MATHEMATICAL DEVELOPMENT
### 4.1. Notation

Let $T$ = time to rupture

$t$ = logarithm of time to rupture

$S$ = stress

$S_0$ = stress for which it is desired to estimate time to rupture

$S_1$ = smaller of two testing stresses to be used

$S_2$ = larger of two testing stresses to be used

$n_1$ = number of tests to be made at $S_1$

$n_2$ = number of tests to be made at $S_2$

$n = n_1 + n_2$

$R = S_1/S_2$

$r = n_1/n_2$

$\alpha, \beta$ = parameters of line relating log $T$ to log $S$, so that

$$\log T = \alpha + \beta \log S \tag{5}$$

$a, b$ = estimates of $\alpha$ and $\beta$

$t_0$ = estimate of logarithm of time to rupture at stress $S_0$

$K$ = total amount of machine time available

$\sigma$ = standard deviation of observations about the line (5), assumed the same for all $S$

$\sigma_b$ = standard deviation of estimated slope

$\sigma_t$ = standard deviation of logarithm of estimated time to rupture

For convenience, take all logarithms to be natural logarithms.

We assume that the parameters of the line are estimated from the data by the usual least-squares procedure applied to the logarithms of the variables. Then, by algebraic manipulation of the usual formulas,[*]

$$\sigma_b^2 = \frac{\sigma^2}{\dfrac{n_1 n_2}{n_1 + n_2}\left(\log \dfrac{S_1}{S_2}\right)^2} \tag{6}$$

$$\sigma_{t_0}^2 = \sigma^2 \left[\frac{n_1 \left(\log \dfrac{S_0}{S_1}\right)^2 + n_2 \left(\log \dfrac{S_0}{S_2}\right)^2}{n_1 n_2 \left(\log \dfrac{S_1}{S_2}\right)^2}\right] \tag{7}$$

The problem is to minimize (6) and (7) by proper choice of $n_1$, $n_2$, $S_1$, and $S_2$, subject to the constraint that a fixed amount of machine time is used. Since $T = e^\alpha S^\beta$, this constraint is

$$n_1 e^\alpha S_1^\beta + n_2 e^\alpha S_2^\beta = K \tag{8}$$

or

$$n_1 S_1^\beta + n_2 S_2^\beta = Ke^{-\alpha} = k \tag{9}$$

It will clarify the solution and yield useful supplementary results to proceed to the final solution in stages, rather than at one step. We shall, therefore, first assume $S_1$ and $S_2$ given and determine the optimum value of $r$; thereafter we shall proceed to determine the optimum value of $R$.

### 4.2. Objective to Get Best Estimate of Slope

The problem is to maximize

$$\frac{\sigma^2}{\sigma_b^2} = \frac{n_1 n_2}{n_1 + n_2}\left(\log \frac{S_1}{S_2}\right)^2 \tag{10}$$

subject to (9), since this will minimize the variance of the slope as given by (6).

Replacing $n_1$ by $rn_2$ in (9) and (10), solving (9) so modified for $n_2$, and substituting the result for $n_2$ in (10) gives

$$\frac{\sigma^2}{\sigma_b^2} = \frac{kr\left(\log \dfrac{S_1}{S_2}\right)^2}{(1 + r)(rS_1^\beta + S_2^\beta)} \tag{11}$$

[*] See FISHER, *Statistical Methods*, Sec. 26, pp. 133, 135. (For full citation, see Sec. 5.)

to be maximized with respect to $r$, without restraint and with $S_1$ and $S_2$ assumed fixed. Setting the derivative with respect to $r$ equal to zero gives

$$\frac{d\left(\dfrac{\sigma^2}{\sigma_b^2}\right)}{dr} = k\left(\log\frac{S_1}{S_2}\right)^2 \frac{(S_2^\beta - r^2 S_1^\beta)}{(1+r)^2(rS_1^\beta + S_2^\beta)^2} = 0 \qquad (12)$$

A solution to (12) is obtained by setting

$$S_2^\beta - r^2 S_1^\beta = 0 \qquad (13)$$

or

$$r = \left(\frac{S_1}{S_2}\right)^{-\frac{\beta}{2}} = R^{-\frac{\beta}{2}} \qquad (14)$$

where we take the positive square root, as $r$ is essentially positive. A check on the second derivative shows that this solution does maximize (11). Substituting (14) into (11) and taking the reciprocal gives

$$\frac{\sigma_b^2}{\sigma^2} = \frac{(S_1^{\frac{\beta}{2}} + S_2^{\frac{\beta}{2}})^2}{k\left(\log\dfrac{S_1}{S_2}\right)^2} \qquad (15)$$

as the minimum ratio of the variance of the estimated slope to the variance of an individual observation, given $S_1$ and $S_2$.

Replacing $S_1$ by $RS_2$ in (15) gives

$$\frac{\sigma_b^2}{\sigma^2} = \frac{(R^{\frac{\beta}{2}} + 1)^2 S_2^\beta}{k(\log R)^2} \qquad (16)$$

Consider $S_2$ as fixed and minimize (16) with respect to $R$; this gives

$$\frac{d\left(\dfrac{\sigma_b^2}{\sigma^2}\right)}{dR} = \frac{2S_2^\beta (R^{\frac{\beta}{2}} + 1)}{kR(\log R)^3}\left[\frac{\beta}{2} R^{\frac{\beta}{2}} \log R - R^{\frac{\beta}{2}} - 1\right] = 0 \qquad (17)$$

whence

$$\frac{\beta}{2}\log R = 1 + R^{-\frac{\beta}{2}} \qquad (18)$$

A check on the second derivative shows that this solution does minimize (16).

Substituting the value for $\log R$ from (18) in (16) gives

$$\frac{\sigma_b^2}{\sigma^2} = \frac{\beta^2 R^\beta S_2^\beta}{4k} \qquad (19)$$

as the minimum value of (16), given $S_2$. Since $\beta$ is negative, it is clear that the larger $S_2$, the smaller the variance; hence there is no finite optimum value of $S_2$. In practice, increases in $S_2$ beyond some point will increase $\sigma^2$, and the optimum value of $S_2$ will be somewhere near this point.[*]

### 4.3. Objective to Get Best Estimate of Time to Rupture at a Given Stress

The problem here is to minimize (7) subject to (9). Replacing $n_1$ by $rn_2$ in (7) and (9), solving (9) so modified for $n_2$, and substituting the result in (7) gives

$$
\frac{\sigma_{t_0}^2}{\sigma^2} = \frac{\left[ r \left( \log \frac{S_0}{S_1} \right)^2 + \left( \log \frac{S_0}{S_2} \right)^2 \right] [r S_1^\beta + S_2^\beta]}{kr \left( \log \frac{S_1}{S_2} \right)^2}
\tag{20}
$$

to be minimized with respect to $r$, without restraint and with $S_1$ and $S_2$ assumed fixed. Setting the derivative of (20) with respect to $r$ equal to zero gives

$$
\frac{d \left( \frac{\sigma_{t_0}^2}{\sigma^2} \right)}{dr} = \frac{1}{kr^2 \left( \log \frac{S_1}{S_2} \right)^2} \left[ r^2 S_1^\beta \left( \log \frac{S_0}{S_1} \right)^2 - S_2^\beta \left( \log \frac{S_0}{S_2} \right)^2 \right] = 0
\tag{21}
$$

From (21) the minimizing solution is given by

$$
r^2 = \left( \frac{S_2}{S_1} \right)^\beta \left( \frac{\log \frac{S_0}{S_2}}{\log \frac{S_0}{S_1}} \right)^2
\tag{22}
$$

or

$$
r = \pm \left( \frac{S_2}{S_1} \right)^{\frac{\beta}{2}} \frac{\log \frac{S_0}{S_2}}{\log \frac{S_0}{S_1}}
\tag{23}
$$

The positive sign is required to make $r$ positive if $S_0$ is less than $S_1$ or greater than $S_2$; the negative sign, if $S_1 < S_0 < S_2$. The second derivative shows that these signs also give minima. Since we are concerned primarily with values of $S_0$ that are below $S_1$, we shall hereafter consider only the positive sign.

[*] It will pay to go somewhat beyond the point at which $\sigma^2$ starts to increase, since the loss from the increase in $\sigma^2$ will for a time be less than the gain from the increase in $S_2$.

Substituting (23) into (20) gives as the minimum ratio of the variance of the estimated rupture time to the variance of an individual observation, with $S_1$ and $S_2$ fixed,

$$\frac{\sigma_{t_0}^2}{\sigma^2} = \frac{\left[ S_1^{\frac{\beta}{2}} \log \frac{S_0}{S_2} + S_2^{\frac{\beta}{2}} \log \frac{S_0}{S_1} \right]^2}{k \left( \log \frac{S_1}{S_2} \right)^2} \tag{24}$$

On replacing $S_1$ by $RS_2$, (24) becomes

$$\frac{\sigma_{t_0}^2}{\sigma^2} = \frac{S_2^{\beta}}{k} \left[ \frac{(1 + R^{\frac{\beta}{2}}) \log \frac{S_0}{S_2}}{\log R} - 1 \right]^2 \tag{25}$$

from which, holding $S_2$ fixed,

$$\frac{d \left( \frac{\sigma_{t_0}^2}{\sigma^2} \right)}{dR} = \frac{2 S_2^{\beta}}{k} \left[ \frac{(1 + R^{\frac{\beta}{2}}) \log \frac{S_0}{S_2}}{\log R} - 1 \right] \left[ \frac{\left( \frac{\beta}{2} R^{\frac{\beta}{2}} \log R - 1 - R^{\frac{\beta}{2}} \right) \log \frac{S_0}{S_2}}{R (\log R)^2} \right]$$

$$= 0 \tag{26}$$

For $S_0$ less than $S_1$, there is no value of $R$ that will make the first bracket equal to zero. Accordingly, the minimizing value of $R$ is given by

$$\frac{\beta}{2} R^{\frac{\beta}{2}} \log R - 1 - R^{\frac{\beta}{2}} = 0 \tag{27}$$

or

$$\frac{\beta}{2} \log R = 1 + R^{-\frac{\beta}{2}} \tag{28}$$

which is the same as (18), the solution for minimizing the variance of the slope. This is, of course, the minimizing value only if $S_2$ is sufficiently large so that the derived $S_1$ is larger than $S_0$.

By differentiating (25) with respect to $S_2$, it can be shown that (25) decreases as $S_2$ increases, so that there is no finite optimum value of $S_2$.

If $S_2$ is chosen greater than $S_0$, while $S_1$ derived from the $R$ satisfying (28) is less than $S_0$, (28) no longer gives the optimum solution. In that case, it turns out that the variance is a monotonically decreasing function of $R$. The best that can be done is to take the largest possible value of $R$, namely, $R = S_0/S_2$, or $S_1 = S_0$. This solution using two testing stresses will not, however, be as good as that putting all the tests at $S_0$.

## 5. REFERENCE

Fisher, R. A.: *Statistical Methods for Research Workers*, 4th and later ed., Oliver and Boyd, Ltd., Edinburgh and London, 1932–1944.

# CHAPTER 10

## EXPERIMENTS IN WHICH THE INDEPENDENT VARIABLE IS KNOWN ONLY IN TERMS OF A PROBABILITY DISTRIBUTION

This chapter was prepared by the editors.

With the exception of the charts and related material, it is based on a memorandum by J. Wolfowitz, Columbia University

# CHAPTER 10

## EXPERIMENTS IN WHICH THE INDEPENDENT VARIABLE IS KNOWN ONLY IN TERMS OF A PROBABILITY DISTRIBUTION

### 1. PROBLEM

### 1.1. Type of Problem

This chapter deals with a type of problem in which it is desired to subject each of a number of items to a test of at least a given severity, but the severity of any one test cannot be measured and cannot be controlled precisely. Thus it is necessary to make an independent study of the probability that a test will attain the specified severity and then to determine the number of tests necessary for each item in order to attain a given probability that at least one test will be of the specified severity. The particular problem discussed is relatively simple, and the chapter is intended more to suggest an interesting and in some respects important type of problem and to indicate an approach to it, than to present a solution or a novel application of statistical techniques.

### 1.2. Specific Problem

A gun barrel is considered fit for service with respect to its pressure-resisting properties if it has fired a round such that the gas pressure developed is not less than a prescribed pressure $K$, $K$ being greater than any pressure that may reasonably occur in service. Special ammunition that develops high pressures is used for proofing the barrels. In order to measure the pressure developed by a round, a hole must be bored in the barrel, which renders it unfit for service. Consequently, it is impossible to ascertain with certainty that a barrel intended for service has actually been proofed. It is, however, possible to draw conclusions about the percentage of barrels proofed from a study of the characteristics of the special ammunition. When a lot of the latter is delivered, a small fixed sample of size $n$ is tested for pressure and the remainder (which is almost the whole lot) is used for proofing, $N$ rounds being fired in each barrel to be proofed. Manufacture of the special ammunition is in a sufficiently satisfactory state that the proportion of defective rounds, that is, those which generate a pressure less than $K$, is usually very small (of the order of 0.001). The problem is, on the basis of the results from the sample of $n$, (1) to set up criteria for accepting or rejecting the lot, and (2) to determine $N$, the number of rounds to be fired in each barrel.
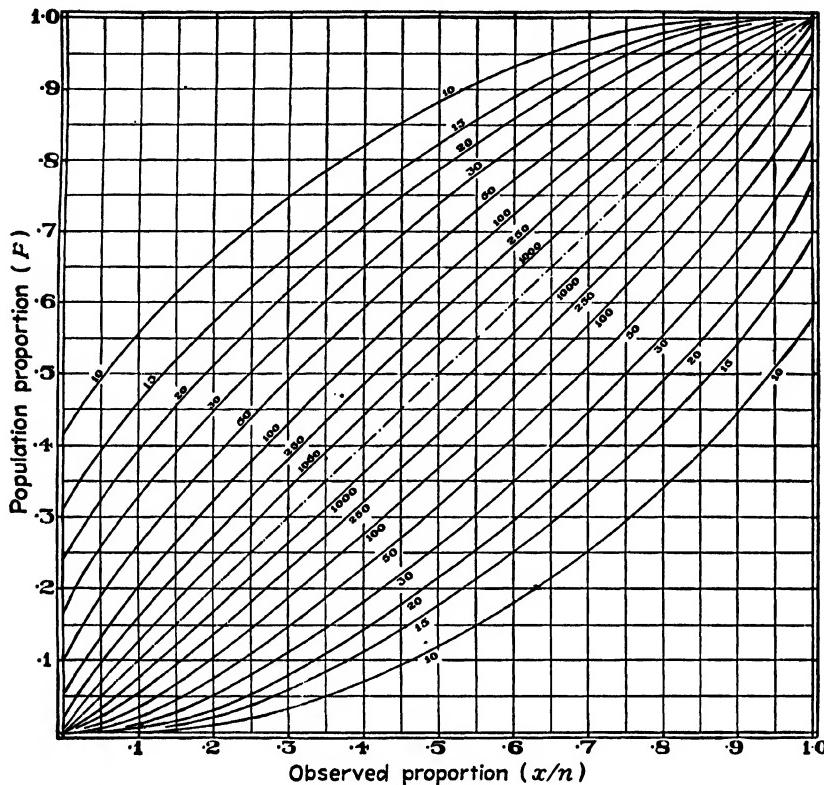
Fig. 10.1.—Ninety-nine percent confidence belts for population proportion.

## 2. SOLUTION

Choose $\gamma$, a confidence coefficient with which to operate, the minimum number of barrels that an acceptable lot should proof, and the minimum tolerable probability of proofing a barrel. From the sample of $n$ obtain an estimate $p_\gamma$ with confidence coefficient $\gamma$ of an upper bound on the proportion of rounds in the lot which, when fired, will generate a pressure less than $K$. How this estimate is to be obtained is described in Sec. 3. Construct a table as follows:

| Number of rounds ($N$) fired in each barrel | With confidence coefficient $\gamma$, the probability that a barrel will be proofed is at least |
|:---:|:---:|
| 1 | $1 - p_\gamma$ |
| 2 | $1 - p_\gamma^2$ |
| . . . . . | . . . . . |
| $i$ | $1 - p_\gamma^i$ |
| . . . . . | . . . . . |

From the table select the smallest $N$ yielding the required probability of proofing a barrel. If this $N$ is too large to enable the lot of ammunition
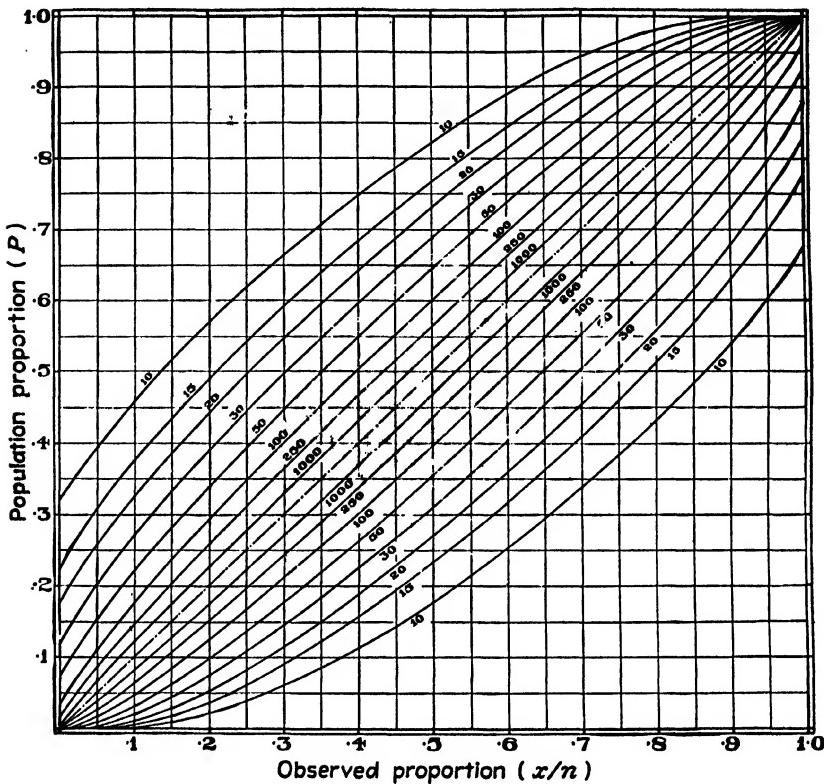
FIG. 10.2.—Ninety-five percent confidence belts for population proportion.

to proof the minimum required number of barrels, reject the lot.    Otherwise, fire $N$ rounds in each barrel.

### 3. CONFIDENCE LIMITS FOR PROPORTION DEFECTIVE

#### 3.1. Attributes

If the upper confidence limit $p_\gamma$ is determined by attributes, that is, by classifying each round in the sample of $n$ as above $K$ or below $K$, it is unnecessary to make any assumption whatever about the distribution of the pressures of the bullets of the lot.    It is only necessary that the sample of $n$ be selected at random.    If it is believed that defective rounds do not occur independently (for example, if a machine functions improperly, all the bullets it produces may be defective and these may be adjacent in the lot), the $N$ bullets to be fired from any gun barrel should be chosen at random from the lot.

Perhaps the most accurate convenient method of finding $p_\gamma$ is with the aid of Thompson's table of percentage points of the incomplete beta function.    Let

Population proportion (P)
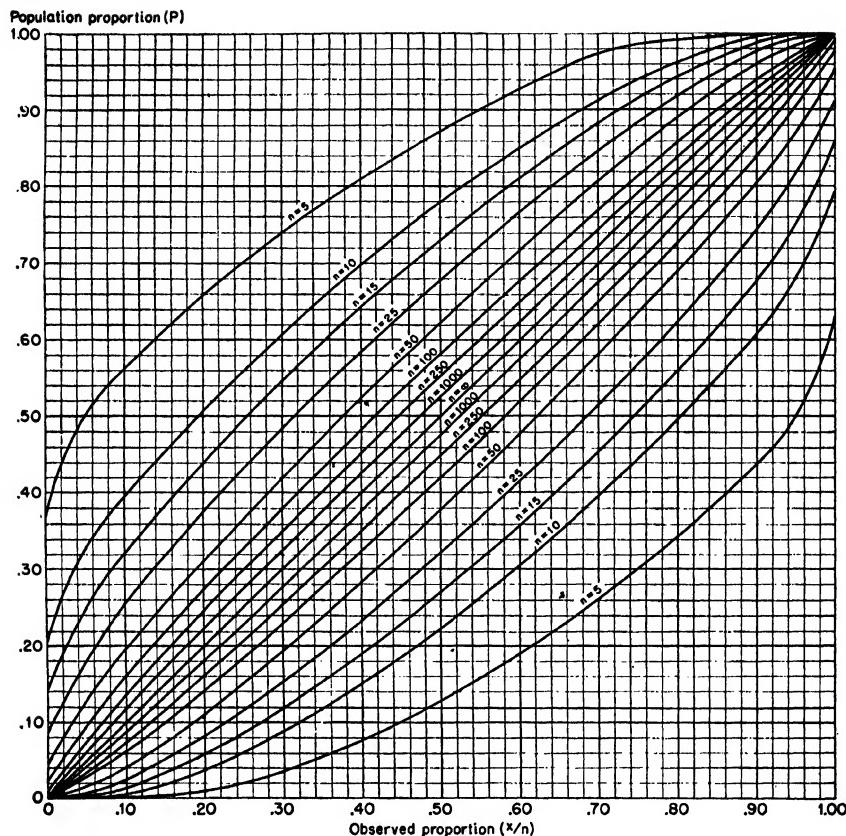


FIG. 10.3.—Ninety percent confidence belts for population proportion.

$$c = \text{number of rounds in the sample of } n$$
$$\text{whose pressures are less than } K \tag{1}$$
$$\alpha = 100(1 - \gamma) \tag{2}$$
$$\nu_1 = 2(c + 1) \tag{3}$$
$$\nu_2 = 2(n - c) \tag{4}$$

Enter the table with $\nu_1$ and $\nu_2$ on the $\alpha$ percentage-point page and find $x$. Then $p_\gamma = 1 - x$. For example, if $\gamma = 0.975$, $c = 0$, $n = 30$, we have $\alpha = 2.5$, $\nu_1 = 2$, $\nu_2 = 60$; from the table we find $x = 0.88430$, or $\hat{p}_{0.975} = 0.1157$. The table of Sec. 2 then appears as follows:

| $N$ | Minimum probability |
|-----|---------------------|
| 1 | 0.8843 |
| 2 | 0.9866 |
| 3 | 0.9985 |

etc.

Population proportion (P)
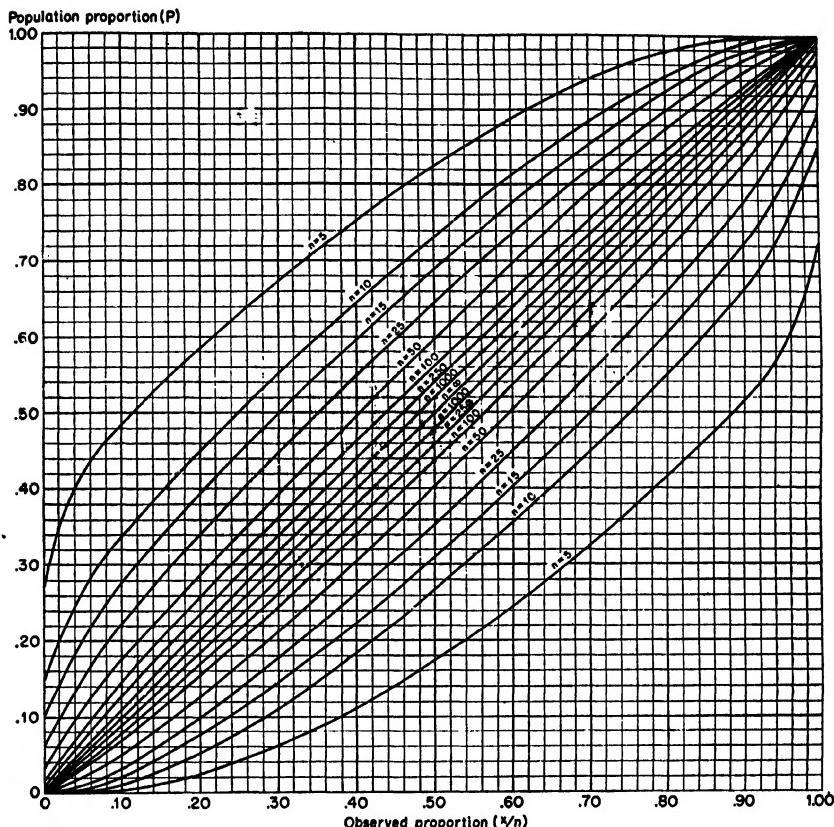


Fig. 10.4.—Eighty percent confidence belts for population proportion.

*Note:* When $c = 0$, $\hat{p}_\gamma = 1 - (1 - \gamma)^{\frac{1}{n}}$.

When $\gamma = 0.995$, $0.975$, $0.95$, or $0.90$, $\hat{p}_\gamma$ may be found from Figs. 10.1, 10.2, 10.3, or 10.4.* These show belts within which we may assert,

* Figures 10.1 and 10.2 are reproduced from Pearson and Clopper, "Confidence Limits," by permission of *Biometrika.* Figures 10.3 and 10.4 are based on computations from Thompson's "Incomplete Beta Function" made (in cooperation with the Statistical Research Group) by R. L. Deckinger of the Bureau of Ordnance, U.S. Navy, and furnished through the courtesy of R. H. Park.   (For full citations, see Sec. 5.)

With regard to Figs. 10.1 and 10.2, Professor E. S. Pearson writes in part: "Actually I believe they are not completely accurate, as the levels at certain points were obtained by interpolation and not by direct calculation from the 'Incomplete Beta Function.' If we reproduce such diagrams ourselves, we shall have to look into this point." (Letter of January 16, 1946, to one of the editors.)

For Figs. 10.3 and 10.4, abscissas were computed for ordinates between 0 and 1 by

with confidence coefficient of, say, $\Gamma$, that the true proportion lies when the observed proportion is $x/n$. For the present purposes $\gamma$ is related to $\Gamma$ by

$$\Gamma = 2\gamma - 1 \qquad (5)$$

Thus, for the example of the preceding paragraph (where $\gamma = 0.975$), $\Gamma = 0.95$. From Fig. 10.2 we see that the upper curve for $n = 30$ cuts the axis $x/n = 0$ at about 0.11 or 0.12.

Still another method of finding $\hat{p}_\gamma$ is by means of the arcsin transformation described in Chap. 16. Thus

$$\hat{p}_\gamma = \sin^2 \left( \arcsin \sqrt{\frac{c}{n}} + \frac{K_{1-\gamma}}{2\sqrt{n}} \right) \qquad (6)$$

where $K_{1-\gamma}$ is the normal deviate exceeded with probability $1 - \gamma$ (see Table 1.1 of Chap. 1). This formula is not satisfactory, however, when $c/n$ is very small, especially not when $c = 0$.

It should be noted that the foregoing assumes the lot to be large enough relative to the sample so that the effect of lot size can be ignored; this will almost always be the case in practice.

### 3.2. Variables

If it can be assumed that the gas pressures of the bullets in the lot are random independent observations on a normally distributed variate, $\hat{p}_\gamma$ can be determined by the method described in Chap. 1, Sec. 8.2 (pp. 42–45).* The value of $\Gamma$ given by equation (5) of this chapter is used as the value of $\gamma$ in Chap. 1; in other words, $L_p$ of Chap. 1 corresponds with $1 - \gamma$ of this chapter.

### 4. SIZE OF PRELIMINARY SAMPLE

While the rounds expended in testing the initial sample of $n$ do not serve directly to proof machine-gun barrels, it would be an error to

---

steps of 0.1, using Thompson's "Incomplete Beta Function." For a confidence coefficient $\Gamma$ (in Fig. 10.3, $\Gamma = 0.90$; in Fig. 10.4, $\Gamma = 0.80$), the upper (lower) abscissa for a given value of $n$ represents the value of $x/n$ that will be exceeded (fallen short of) with probability $(1 - \Gamma)/2$ in samples of size $n$ from a binomial population with parameter $P$. A smooth curve was passed through the 11 points, but is shown on the chart only over the range $1/n' \leq x/n \leq 1 - 1/n'$, where $n'$ is the next sample size above $n$ shown on the chart. Outside this range the curve was bent so that at abscissas 0 and 1 the ordinates are $1 - (1 - \Gamma)^{1/n}$ and $(1 - \Gamma)^{1/n}$, respectively, instead of $1 - [(1 - \Gamma)/2]^{1/n}$ and $[(1 - \Gamma)/2]^{1/n}$. The reason for the difference in computation at $x/n = 0$ and 1 is that there the confidence interval must lie entirely to one side of the observed proportion. In deciding where to bend the curve, it was assumed that a given curve will never be used in the intervals $0 < x/n < 1/n'$ and $1 - 1/n' < x < 1$, and that only for interpolation will it ever be used in the intervals $1/n' < x/n < 1/n$ and $1 - 1/n < x/n < 1 - 1/n'$.

\* A slightly more accurate method is given in Chap. 1, Sec. 11.4.6, Problem II, where $\epsilon$ corresponds with $1 - \gamma$ of this chapter (see pp. 76–77).

characterize them as wasted and to try to keep $n$ as small as possible. Actually these rounds supply information about the quality of the ammunition, and more information can easily result in an actual saving of ammunition.   Suppose, for example, that a lot contains 10,000 rounds, of which (unknown to us) 10 (0.1 percent) are defective, that we wish to operate with a 0.99 confidence coefficient, and that the minimum tolerable probability of proofing a barrel is to be 0.99.   Suppose further that in a sample of size 20 ($n = 20$) no defectives appear ($c = 0$).   Proceeding as in Sec. 3.1, we construct the following table:

$$n = 20 \qquad \gamma = 0.99 \qquad c = 0$$
$$\alpha = 0.01 \qquad \nu_1 = 2 \qquad \nu_2 = 40$$
$$x = 0.79433 \qquad \hat{p}_{0.99} = 0.20567$$

| $N$ | Minimum probability |
|-----|---------------------|
| 1   | 0 79                |
| 2   | 0 96                |
| 3   | 0.99+               |

Three rounds are needed for each barrel, so the total number of barrels proofed by the lot will be

$$\frac{10,000 - 20}{3} = 3,326 \qquad (7)$$

Now, suppose that a sample of 60 ($n = 60$) had been taken and no defectives found.   Then

$$n = 60 \qquad \gamma = 0.99 \qquad c = 0$$
$$\alpha = 0.01 \qquad \nu_1 = 2 \qquad \nu_2 = 120$$
$$x = 0.92612 \qquad \hat{p}_{0.99} = 0.07388$$

| $N$ | Minimum probability |
|-----|---------------------|
| 1   | 0.93                |
| 2   | 0.99+               |

Two rounds suffice for each barrel, and the total number of barrels proofed by the lot would be

$$\frac{10,000 - 60}{2} = 4,970 \qquad (8)$$

a 49 percent increase over the previous result.

Of course the number of defectives in the sample of $n$ is a chance variable whose distribution depends on $n$, among other parameters.   On the basis of previous experience with the ammunition being used, it is possible to determine that value of $n$ which is optimum in the sense that if the quality of the ammunition is essentially unchanged with time, the

average number of machine-gun barrels successfully proofed will be maximized.

## 5. REFERENCES

CLOPPER, C. J., and PEARSON, E. S.: "The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial," *Biometrika*, Vol. 26 (1934), pp. 404–413.

THOMPSON, CATHERINE M.: "Percentage Points of the Incomplete Beta Function," *Biometrika*, Vol. 32 (1941), pp. 168–181.

# CHAPTER 11

# PLANNING AN EXPERIMENT FOR ESTIMATING THE MEAN AND STANDARD DEVIATION OF A NORMAL DISTRIBUTION FROM OBSERVATIONS ON THE CUMULATIVE DISTRIBUTION

This chapter was prepared by
Milton Friedman, University of Chicago

# CHAPTER 11

## PLANNING AN EXPERIMENT FOR ESTIMATING THE MEAN AND STANDARD DEVIATION OF A NORMAL DISTRIBUTION FROM OBSERVATIONS ON THE CUMULATIVE DISTRIBUTION

### 1. THE PROBLEM

A class of statistical problems of frequent occurrence arises from situations in which a characteristic is conceived to be normally distributed among the individuals of a population, but measurement of any one individual can show only whether the characteristic is above or below a certain level. The most familiar form of this problem concerns the lethality of varying doses of drugs. The electric impulse necessary to ignite a charge of powder in a certain system represents another such problem; for, although in this case it is possible to apply increasing impulses until ignition occurs, there is some evidence that an impulse that fails to cause ignition may alter the sensitivity of the system. In such situations a single stimulus of known intensity is applied to each individual. The resulting data show the proportions of the individuals tested at various intensities whose sensitivity fell *below* those intensities; in other words, the data for each level of stimulus intensity provide one observation on the cumulated distribution of sensitivities. These observations may be analyzed to estimate the mean and standard deviation of the sensitivity of the individuals comprising the population.

The present chapter is concerned with certain aspects of the planning of experiments that yield observations on the cumulated distribution and are analyzed to determine the mean and standard deviation of the distribution. For concreteness, the discussion is given in terms of the specific problem of planning an experiment to study the sensitivity of a radio proximity fuze. A radio proximity fuze, as its name suggests, makes use of radio waves to explode a projectile when and if it comes into proximity with a target; unlike time fuzes or point-detonating fuzes, it requires neither that a definite time of flight be set in advance nor that the target be hit directly. If proximity fuzes of a given model behaved identically there would, for fixed conditions, be a definite distance from the target within which all fuzes would operate and beyond which none would operate. Actually, there are random variations in the fuzes and in the impulses returned by the target; so if a number of fuzes pass exactly the same distance from a target under the same conditions, some operate

341

and some do not. An important parameter of fuze performance is, therefore, the probability that a fuze will operate if it passes within a certain distance of the target.

## 2. METHOD OF ANALYSIS

It has been determined from a certain set of 1,211 test firings made under relatively homogeneous conditions that the probability that a nondefective fuze will operate if it passes within a certain distance of the target can be well represented by assuming (a) that each nondefective fuze is characterized by a definite sensitivity, or maximum distance at which it will function, and (b) that these sensitivities are normally distributed among fuzes.

Let $S$ be the sensitivity of a fuze, defined as the maximum distance at which that fuze fired under the given circumstances will function

$\mu$ be the true (population) arithmetic mean sensitivity

$\sigma$ be the true (population) standard deviation of sensitivity

Then under the above assumptions the probability that a fuze will have a sensitivity between $S$ and $S + dS$ is

$$\frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(S-\mu)^2}{2\sigma^2}} dS \qquad (1)$$

The proportion $T(d)$ of fuzes passing a given distance $d$ from the target that operate is then the integral of (1) from $d$ to $\infty$, since every fuze whose sensitivity is greater than $d$ will operate at $d$, while every fuze whose sensitivity is less will not. Therefore,

$$T(d) = \frac{1}{\sigma \sqrt{2\pi}} \int_d^\infty e^{-\frac{(S-\mu)^2}{2\sigma^2}} dS \qquad (2)$$

In firing trials, it is not possible to observe the sensitivity of a fuze. Each fuze can be fired only once; all that can be observed is the minimum distance from the target and whether the fuze operated in proximity to the target.* The observations obtained are therefore observations on $T(d)$, the integral of the normal curve, and not directly on $S$.

Methods for analyzing such data have been developed, primarily in connection with dosage-mortality curves.† These methods involve writ-

---

* Actually, the entire trajectory of the fuze can be observed and also the actual point where the fuze operated. However, as these play no role in the problem considered here, they are ignored.

† See BLISS, "The Calculation of the Dosage-mortality Curve," "The Comparison of Dosage-mortality Data," "Estimating the Dosage-mortality Curve," and "The Calculation of the Time Mortality Curve"; also FISHER and YATES, *Tables*, pp. 8–12. (For full citations, see Sec. 6.)

ing (2) as

$$T(d) = \frac{1}{\sqrt{2\pi}} \int_Y^\infty e^{-\frac{x^2}{2}} dx \tag{3}$$

where

$$Y = \frac{d - \mu}{\sigma} \tag{4}$$

An observed value of $T(d)$ can be converted into an estimate of $Y$ by means of tables of the normal probability integral.* The estimate of $Y$ may be denoted by $y$. Now, $Y$ is linearly related to $d$ by

$$Y = -\frac{\mu}{\sigma} + \frac{1}{\sigma} d \tag{5}$$

Hence the observed data may be analyzed by fitting a straight line to the observed values $(y,d)$. The usual direct methods of fitting a straight line cannot be used, however, since the variance of $y$ is not a constant but depends on $Y$; instead, a method of successive approximations is used. Tables are available to facilitate fitting the straight line by this method.† To avoid negative values, some of these tables use "probits," defined as normal deviates increased by 5, instead of the normal deviates themselves.

The function (5) was fitted to the data observed in the 1,211 test firings on fuze performance mentioned previously. The estimates of $\mu$ and $\sigma$ turned out to be $A$ and $B$, respectively.‡ Figure 11.1 shows the fitted function (5), as well as 90 percent confidence limits on it. These confidence limits were computed from the equation§

$$\sigma_Y^2 = s_y^2 \left[ \frac{1}{\sum\limits_{i=1}^{N} w_i} + (d_i - \bar{d})^2 \frac{1}{\sum\limits_{i=1}^{N} w_i(d_i - \bar{d})^2} \right] \tag{6}$$

where $N$ = number of observed points to which the straight line was fitted

$w_i$ = weight given to the $i$th point in getting the last approximation

$\bar{d}$ = weighted average of observed values of $d$, that is,

$$\sum_{i=1}^{N} w_i d_i \Big/ \sum_{i=1}^{N} w_i$$

* For example, Table 1.5 in Chap. 1 of this volume.

† For example, FISHER and YATES, *Tables*, p. 49. Another useful set of tables is contained in Brown, *Tables to Facilitate the Analysis of Sensitivity Data*. This shows the ordinate $Z$ of a normal curve, the probability $P$ that a given normal deviate will be exceeded, $Q = 1 - P$, $Z/PQ$, and $Z^2/PQ$, all to 5 decimal places for values of $Y$ from 0 to 3 by steps of 0.01.

‡ The numerical values of $A$ and $B$, which cannot be published, are immaterial for the present purposes.

§ See BLISS, "The Calculation of the Dosage-mortality Curve."

$s^2$ = estimated variance of an observation about the last approximation, that is, if $Y_i$ is the regression value for the $i$th point computed from the last approximation, then

$$s_y^2 = \frac{\sum_{i=1}^{N} w_i(y_i - Y_i)^2}{N - 2} \qquad (7)$$

The problem then arose of collecting similar data under different conditions. Theoretical considerations gave little indication of the way in which the performance of the fuze would be affected by the difference
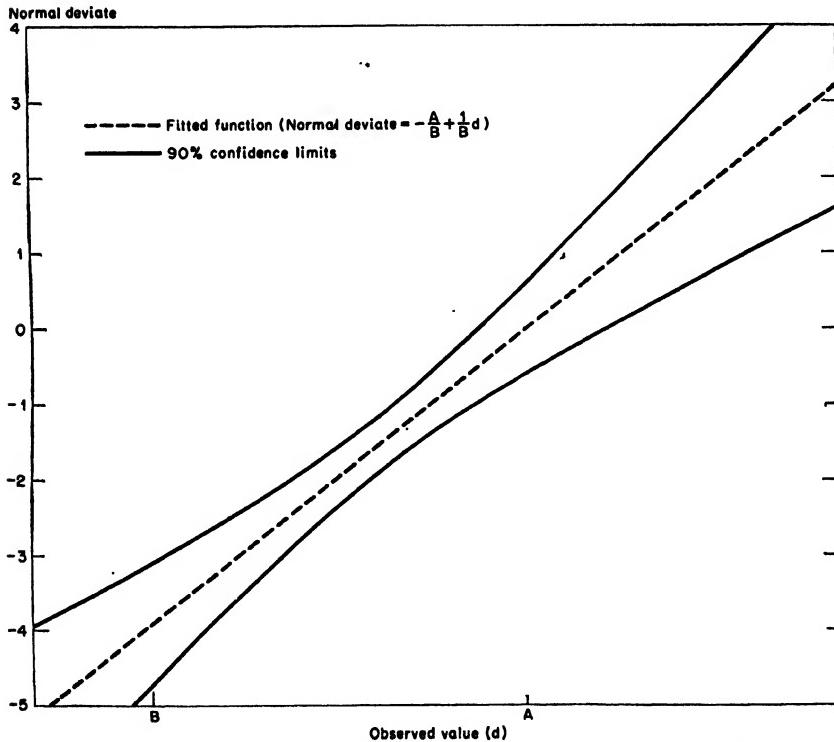


FIG. 11.1.—Fitted function with 90 percent confidence limits based on 1211 observations.

in conditions. The scale of the experiment had to be limited to at most 250 rounds. The statistical problem, with which the rest of this chapter deals, is that of distributing these rounds among various possible distances from the target in such a way as to get as accurate an estimate as possible of $T(d)$. This problem serves to illustrate a number of considerations

pertinent to such experiments; while no general, theoretical solution is offered, the mode of approach should be at least suggestive in similar situations.

### 3. STATISTICAL CONSIDERATIONS AFFECTING THE PLANNING OF THE EXPERIMENT

A paradox frequently arises in planning experiments. The most efficient design depends on the true value of the parameter under investigation, but if this were known there would be no need for the experiment. This paradox is usually evaded by basing the experimental design on some advance estimate of the parameters and trying to adjust the design so that the efficiency of the experiment is not much affected by minor differences between what is true and what is assumed. This method itself raises difficulties, for example, how much of the efficiency that the optimum design would have if the advance estimate were correct should be sacrificed to gain efficiency if the advance estimate is not correct?

This difficulty is present in marked degree in the problem under consideration and is aggravated by another difficulty, namely, that two parameters are to be estimated and the experimental design that is best for estimating one is not best for estimating the other. The two parameters to be estimated are $\mu$ and $\sigma$, or, what is equivalent, the slope and intercept of (5). The experiment would be most efficient if it could minimize both the variance of the estimate of $\mu$ and that of the estimate of $\sigma$; unfortunately, the distribution of observations that minimizes the variance of the estimate of $\mu$ is decidedly different from that which minimizes the variance of the estimate of $\sigma$.

If $\sigma$ were known, the variance of the estimate of $\mu$ would be minimized by taking all observations at a single distance equal to $\mu$. But since $\sigma$ is not known, it would be self-defeating to take all observations at a single distance, because (a) this would make it impossible to correct the advance estimate of $\mu$, and (b) it would give no information at all about $\sigma$. If the observed value of $T(d)$ were 50 percent, the estimate of $\mu$ from the experiment would be equal to the value of $d$ at which the firings were made, that is, it would be the same as the advance estimate. If the observed value of $T(d)$ were other than 50 percent, the experiment would indicate that the advance estimate of $\mu$ was wrong but no new estimate could be made. Only one point on the straight line (5) would be available, so its slope could not be estimated; hence, it would be impossible to estimate the value of $d$ at which $T(d) = 50$ percent. These statements are equivalent, of course, to saying that no estimate of $\sigma$ would be possible either. The most that could be learned from an experiment with all observations at an estimated $\mu$ would be whether this estimate was correct.

The variance of the estimate of $\sigma$ would be minimized by taking all observations at two values of $d$, approximately 1.6 times the estimated

value of $\sigma$ on either side of the estimated value of $\mu$. Even though this would not give as accurate an estimate of $\mu$ as could be obtained by placing the observations closer to $\mu$, it is a workable arrangement in that it would yield estimates of both $\mu$ and $\sigma$ even though the advance estimates were in error.

From these considerations alone, the appropriate compromise would appear to be to place the observations in two groups located symmetrically about the estimated $\mu$ but closer to it than the distance that would minimize the variance of the estimate of $\sigma$. To decide how much closer would require deciding how much it is worth sacrificing accuracy in $\sigma$ to gain accuracy in $\mu$.

Other considerations must, however, be taken into account. Placing all the observations at two points puts great reliance on the validity of equation (2) as a description of $T(d)$. This relation was derived from a single body of data, for which its validity cannot be considered conclusively established; it may, therefore, not be valid for the new experiment. Observations at only two values of $d$ cannot, however, show whether it is valid or not, since there will be only two points to which to fit (5), and a straight line can always be fitted perfectly to two points. In order to determine whether (2), and hence (5), describe the data reasonably well, it is necessary to spread the observations among more values of $d$. The desirability of spreading the observations was reinforced in the present experiment by the fact that the data were to be used for other purposes, in particular, to study the location of points of functioning, for which a wide spread seemed desirable.

Other considerations to be taken into account in the particular experiment were (1) it was not feasible to use distances greater than about $A + \frac{5}{3}B$; (2) most of the firings in the earlier experiment had been at short distances, so that it seemed desirable to have some firings in the new experiment at short distances in order to make a direct comparison of the results of the two experiments, independent of any assumptions about the form of $T(d)$.

### 4. POSSIBLE DESIGNS

In the light of the preceding considerations, five alternative distributions of 250 observations were considered. These are shown in Table 11.1.

Plans 1 and 2 are suggested by considering how to minimize the variance of the estimate of $\sigma$. The advance estimates, based on the previous experiment, were $A$ for $\mu$ and $B$ for $\sigma$. The 250 observations in Plans 1 and 2 are divided into two groups and placed symmetrically on either side of the mean. In Plan 2, they are about as far apart as is required to minimize the estimate of $\sigma$; in Plan 1, they are brought closer

together to get a more accurate estimate of $\mu$.    Plan 3 spreads the observation evenly over the five classes spanned by Plan 2; it is symmetrical, and its largest distance is the largest that is feasible.    Plans 4 and 5 are attempts to get some data for the short distances for comparison with the previous experiment.    In Plan 4, the observations for the low distances are obtained by omitting those at $A$, since if $A$ is the true mean, such firings give no information about $\sigma$.    In Plan 5, the observations for the low distances are taken equally from the five groups in Plan 3.

To measure the relative efficiency of these five possible designs, 90 percent confidence limits on (5) were computed for each, assuming that the correct straight line was that obtained in the earlier experiment,

$$Y = -\frac{A}{B} + \frac{1}{B}\,d \tag{8}$$

The confidence limits for Plans 1 and 2 are plotted in Fig. 11.2, together with the confidence limits obtained from the earlier experiment (shown also in Fig. 11.1).[*]    Table 11.2 shows these confidence limits and the assumed correct values in somewhat different form, namely, in terms of $100\,T(d)$, the percentage of nondefective fuzes.

TABLE 11.1
ALTERNATIVE DISTRIBUTIONS OF 250 OBSERVATIONS

| Distance | Number of rounds to be fired under plan | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| $A - 2\frac{3}{8}B$ | | | | 20 | 20 |
| $A - 1\frac{5}{8}B$ | | | | 30 | 30 |
| $A - 1\frac{3}{8}B$ | | 125 | 50 | 50 | 40 |
| $A - \frac{5}{8}B$ | 125 | | 50 | 50 | 40 |
| $A$ | | | 50 | | 40 |
| $A + \frac{5}{8}B$ | 125 | | 50 | 50 | 40 |
| $A + 1\frac{3}{8}B$ | | 125 | 50 | 50 | 40 |

Two points are fairly clear from Fig. 11.2 and Table 11.2    (1) Any of the five plans for 250 observations gives as much information as, if not somewhat more than, the 1,211 observations used in the earlier experiment; (2) the five possible plans do not differ substantially in efficiency. Plan 1, which emphasizes most strongly the accuracy of the estimate of $\mu$, gives therefore the narrowest confidence limits for distances near the mean.    Plan 2, which emphasizes most the accuracy of the estimate of $\sigma$, gives the narrowest limits for distances far from the mean but the widest

[*] The confidence limits for Plans 3, 4, and 5 are omitted from Fig. 11.2 only because they are so close to those for Plans 1 and 2 that if all five were plotted on a figure the size of Fig. 11.2, it would be impossible to distinguish the five lines.

TABLE 11.2

NINETY PERCENT CONFIDENCE LIMITS ON ASSUMED FUNCTION FOR FIVE ALTERNATIVE PLANS FOR 250 OBSERVATIONS

(In percent of nondefective fuzes operating)

| Distance, $d$ | Assumed function | 90% confidence limits for plan | | | | | | | | | | 90% confidence limits from earlier data | |
| | | 1 | | 2 | | 3 | | 4 | | 5 | | | |
| | | Lower | Upper | Lower | Upper | Lower | Upper | Lower | Upper | Lower | Upper | Lower | Upper |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $A - 2\frac{5}{8}B$ | 100.0 | 99.7 | 100.0 | 99.9 | 100.0 | 99.8 | 100.0 | 99.8 | 100.0 | 99.7 | 100.0 | 99.9 | 100.0 |
| $A - 2\frac{1}{8}B$ | 99.9 | 98.6 | 100.0 | 99.2 | 100.0 | 98.8 | 100.0 | 99.0 | 100.0 | 98.7 | 100.0 | 99.6 | 100.0 |
| $A - 1\frac{5}{8}B$ | 99.3 | 94.9 | 99.9 | 96.1 | 99.9 | 95.3 | 99.9 | 95.6 | 99.9 | 95.1 | 99.9 | 97.7 | 99.8 |
| $A - 1\frac{1}{8}B$ | 95.0 | 85.7 | 98.7 | 86.5 | 98.5 | 86.1 | 98.6 | 86.2 | 98.6 | 85.6 | 98.7 | 90.2 | 97.7 |
| $A - \frac{5}{8}B$ | 79.7 | 68.0 | 88.4 | 65.8 | 89.5 | 67.8 | 88.5 | 66.9 | 89.0 | 67.0 | 88.9 | 65.8 | 89.5 |
| $A$ | 50.8 | 40.7 | 60.9 | 36.1 | 65.4 | 39.8 | 61.8 | 37.8 | 63.7 | 38.7 | 62.9 | 27.7 | 73.7 |
| $A + \frac{5}{8}B$ | 21.5 | 12.5 | 33.4 | 11.4 | 35.4 | 12.4 | 33.6 | 11.6 | 35.1 | 11.6 | 35.0 | 5.1 | 52.2 |
| $A + 1\frac{1}{8}B$ | 5.5 | 1.5 | 15.2 | 1.7 | 14.1 | 1.6 | 14.7 | 1.5 | 15.2 | 1.4 | 15.8 | 0.4 | 30.5 |
| $A + 1\frac{5}{8}B$ | 0.8 | 0.1 | 5.4 | 0.1 | 4.1 | 0.1 | 5.0 | 0.1 | 5.0 | 0.1 | 5.6 | 0.0 | 14.3 |
| $A + 2\frac{1}{8}B$ | 0.1 | 0.0 | 1.5 | 0.0 | 0.8 | 0.0 | 1.3 | 0.0 | 1.2 | 0.0 | 1.5 | 0.0 | 5.2 |

limits for distances near the mean. The other plans are in general intermediate.

On the basis of this evidence, it was decided to recommend Plan 5. This plan covers the largest number of values of $d$ and therefore comes closest to satisfying objectives other than the minimization of the variance of estimates of $\mu$ and $\sigma$; at the same time, it does not sacrifice much efficiency in the estimation of $\mu$ and of $\sigma$.



Normal deviate

Observed value (d)
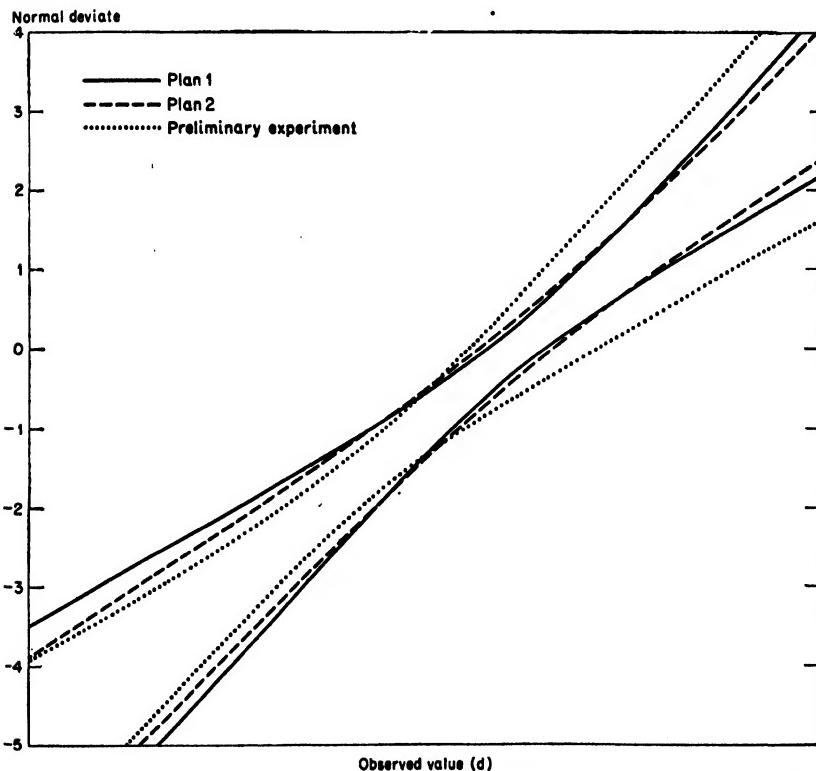
FIG. 11.2.—Ninety percent confidence limits on assumed function for two experimental designs utilizing 250 observations, and for preliminary experiment based on 1211 observations.

## 5. TECHNICAL NOTES*

The problem of determining the optimum distribution of observations for estimating $\mu$ and $\sigma$ is considerably simplified by the following two requirements, which seem plausible and will be accepted here without rigorous proof:

* This section presumes a knowledge of the methods of analyzing probit data (see References, Sec. 6).

(1) The symmetry of the normal distribution implies that the optimum distribution of observations will be symmetrical about the true mean $\mu$. Note that replacing $d - \mu$ by its negative will have no effect on variances, since these involve only (a) the weights, which depend only on the absolute value of $(d - \mu)/\sigma$, and (b) the squares of quantities like $d - \mu$.

(2) Since the problem is one of fitting a straight line containing two parameters, the optimum distribution will involve placing all the observations at two values of $d$. These two values of $d$ enter into the variances in the form of linear combinations of weighted squared deviations. If one pair of values, $d_1$ and $d_2$, such that $d_1 - \mu = -(d_2 - \mu)$, contributes more to the minimization of the variance than another pair of values, say $d_3$ and $d_4$, such that $d_3 - \mu = -(d_4 - \mu)$, it will continue to do so no matter how many observations are shifted from $d_3$ and $d_4$ to $d_1$ and $d_2$.

These two assumptions reduce the problem to one of determining a pair of values $d_1$ and $d_2$ such that $d_1 - \mu = -(d_2 - \mu)$, at each of which half the observations are to be placed.

Let $\mu$ = population mean

$\quad \sigma$ = population standard deviation

$\quad m$ = estimate of $\mu$

$\quad s$ = estimate of $\sigma$

$\quad Y_i = \dfrac{d_i - \mu}{\sigma} \ (i = 1, 2)$

$\quad y_i = \dfrac{d_i - m}{s}$ = estimate of $Y_i$

$\quad T_i = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{Y_i}^{\infty} e^{-\frac{x^2}{2}} \, dx$ = true proportion above $d_i$

$\quad t_i$ = estimate of $T_i$

Observed values of $t_i$ give, from tables of probits or of the normal distribution, estimates of $y_i$, and hence two equations

$$y_1 = \frac{d_1 - m}{s} \tag{9}$$

$$y_2 = \frac{d_2 - m}{s} \tag{10}$$

Solving for $m$ and $s$ gives

$$m = \frac{y_1 d_2 - y_2 d_1}{y_1 - y_2} \tag{11}$$

$$s = \frac{d_1 - d_2}{y_1 - y_2} \tag{12}$$

It is known* that

$$\sigma_{y_i}^2 = \frac{T_i(1 - T_i)}{n_i z_i^2} \tag{13}$$

where

\* See Bliss, "The Calculation of the Dosage-mortality Curve."

$$n_i = \text{number of observations at } d_i$$

$$z_i = \frac{1}{\sqrt{2\pi}} e^{-\frac{Y_i^2}{2}}$$

To a first approximation,

$$\sigma_m^2 = \left(\frac{\partial m}{\partial y_1}\right)_{y_i=Y_i}^2 \sigma_{y_1}^2 + \left(\frac{\partial m}{\partial y_2}\right)_{y_i=Y_i}^2 \sigma_{y_2}^2 \tag{14}$$

$$\sigma_s^2 = \left(\frac{\partial s}{\partial y_1}\right)_{y_i=Y_i}^2 \sigma_{y_1}^2 + \left(\frac{\partial s}{\partial y_2}\right)_{y_i=Y_i}^2 \sigma_{y_2}^2 \tag{15}$$

The partial derivatives in (14) and (15) may be evaluated as follows:

$$\frac{\partial m}{\partial y_1} = \frac{y_2(d_1 - d_2)}{(y_1 - y_2)^2} \tag{16}$$

$$\frac{\partial m}{\partial y_2} = \frac{-y_1(d_1 - d_2)}{(y_1 - y_2)^2} \tag{17}$$

$$\frac{\partial s}{\partial y_1} = \frac{-(d_1 - d_2)}{(y_1 - y_2)^2} \tag{18}$$

$$\frac{\partial s}{\partial y_2} = \frac{d_1 - d_2}{(y_1 - y_2)^2} \tag{19}$$

Hence

$$\sigma_m^2 = \frac{Y_2^2(d_1 - d_2)^2}{(Y_1 - Y_2)^4} \frac{T_1(1 - T_1)}{n_1 z_1^2} + \frac{Y_1^2(d_1 - d_2)^2}{(Y_1 - Y_2)^4} \frac{T_2(1 - T_2)}{n_2 z_2^2} \tag{20}$$

$$\sigma_s^2 = \frac{(d_1 - d_2)^2}{(Y_1 - Y_2)^4} \frac{T_1(1 - T_1)}{n_1 z_1^2} + \frac{(d_1 - d_2)^2}{(Y_1 - Y_2)^4} \frac{T_2(1 - T_2)}{n_2 z_2^2} \tag{21}$$

Substituting for $Y_1$ and $Y_2$ in (20) and (21) and simplifying gives

$$\sigma_m^2 = \frac{\sigma^2}{(d_1 - d_2)^2} \left[ (d_2 - \mu)^2 \frac{T_1(1 - T_1)}{n_1 z_1^2} + (d_1 - \mu)^2 \frac{T_2(1 - T_2)}{n_2 z_2^2} \right] \tag{22}$$

$$\sigma_s^2 = \frac{\sigma^4}{(d_1 - d_2)^2} \left[ \frac{T_1(1 - T_1)}{n_1 z_1^2} + \frac{T_2(1 - T_2)}{n_2 z_2^2} \right] \tag{23}$$

So far, no use has been made of the symmetry requirement. Equations (22) and (23) are therefore valid for any pair of values of $d_1$ and $d_2$. Symmetry implies that

$$n_1 = n_2 \tag{24}$$

$$(d_1 - \mu) = -(d_2 - \mu) \tag{25}$$

$$T_1 = 1 - T_2 \tag{26}$$

$$z_1 = z_2 \tag{27}$$

From (25),

$$d_1 - d_2 = 2(d_1 - \mu) \tag{28}$$

Substituting into (22) and (23) gives

$$\sigma_m^2 = \frac{\sigma^2}{2} \frac{T_1(1 - T_1)}{n_1 z_1^2} \tag{29}$$

$$\sigma_s^2 = \frac{\sigma^4}{2(d_1 - \mu)^2} \frac{T_1(1 - T_1)}{n_1 z_1^2} = \frac{\sigma^2}{2Y_1^2} \frac{T_1(1 - T_1)}{n_1 z_1^2} \tag{30}$$

$\sigma_m^2$ will be a minimum if $T_1(1 - T_1)/z_1^2$ is a minimum with respect to $d_1$ or, what is the same thing, with respect to $Y_1$. But $T_1(1 - T_1)/z_1^2$ is a minimum at $Y_1 = 0$, as can be demonstrated by differentiating or, more simply, by a glance at the tabled values of $z_1^2/T_1(1 - T_1)$.* Hence, to minimize the variance of $m$, with $\sigma$ assumed known, it would be necessary to put all the observations at $\mu$.

$\sigma_s^2$ will be a minimum if $T_1(1 - T_1)/z_1^2 Y_1^2$ is a minimum. Differentiating yields an equation that would have to be solved numerically. Since, as noted above, $z_1^2/[T_1(1 - T_1)]$ is tabled, it is simpler to compute directly $Y_1^2 z_1^2/[T_1(1 - T_1)]$ and determine numerically the value of $Y_1$ at which it reaches a maximum. This gives approximately 1.574 or 1.575 as the value of $Y$ for which $\sigma_s^2$ is a minimum.† Hence, the variance of $s$ will be minimized by putting half the observations at $d_1 = \mu + 1.57\sigma$, and half at $d_2 = \mu - 1.57\sigma$.

### 6. REFERENCES

BLISS, C. I.: "The Calculation of the Dosage-mortality Curve," *Annals of Applied Biology*, Vol. 22 (1935), pp. 134–167.

———: "The Comparison of Dosage-mortality Data," *Annals of Applied Biology*, Vol. 22 (1935), pp. 307–333.

———: "Estimating the Dosage-mortality Curve," *Journal of Economic Entomology*, Vol. 28 (1935).

———: "The Calculation of the Time Mortality Curve," *Annals of Applied Biology*, Vol. 24 (1937), pp. 815–852.

BROWN, RICHARD H.: *Tables to Facilitate the Analysis of Sensitivity Data*, NavOrd Report No. 205–45, Bureau of Ordnance, Navy Department, Washington, D.C., September 5, 1945, 15 pp.

FISHER, RONALD A., and YATES, FRANK: *Statistical Tables for Biological, Agricultural and Medical Research*, 2d ed., Oliver and Boyd, Ltd., London and Edinburgh, 1943, 98 + viii pp.

* See Fisher and Yates, *Tables*, Table XI, where what is designated here $z_1^2/[T_1(1 - T_1)]$ is designated $Z^2/PQ$. A table for finer intervals of $Y$ is available from the Division of Pharmacology, Food and Drug Administration, United States Department of Agriculture, Washington 25, D.C.

† The computed values of $Y_1^2 z_1^2/[T_1(1 - T_1)]$ for a few values in the neighborhood of the maximum are, when diminished by 0.608 and multiplied by $10^5$

| $Y_1$ | 1.55 | 1.56 | 1.57 | 1.58 | 1.59 |
|---|---|---|---|---|---|
| $\dfrac{Y_1^2 z_1^2}{T_1(1 - T_1)}$ | 7 | 30 | 41 | 40 | 29 |

# CHAPTER 12

# ABANDONING AN EXPERIMENT PRIOR TO COMPLETION

This chapter was prepared by the editors.

The idea is due to L. J. Savage, University of Chicago, and the techniques are due to Kenneth J. Arnold, University of Wisconsin, and M. A. Girshick, Stanford University

# CHAPTER 12

## ABANDONING AN EXPERIMENT PRIOR TO COMPLETION

### 1. INTRODUCTION

Every experiment is designed in the light of certain presumptions about the underlying phenomena to be studied, if only the presumption that the phenomena warrant study. It is not always possible, however, to confirm these presumptions in advance. In such cases it may be desirable to introduce a sequential element into the design of the experiment so that the experimental program can be adjusted to any facts about the basic assumptions disclosed during the course of the investigation. The present chapter considers a simple design of this type, where an experiment is to be undertaken to estimate a parameter of a continuous distribution, but provision is to be made to abandon the experiment prior to completion if the accumulated evidence at any stage reveals that the underlying population does not warrant study.

### 1.1. Nature of the Problem

The circumstances giving rise to the need of an experiment that provides for possible abandonment are these: A new mechanism has been constructed whose performance, in functioning models, is thought to be superior to that of similar standard mechanisms. It is accordingly desired to determine the average and variability of a particular performance characteristic from an experiment of sufficient size to give reliable estimates of these parameters. There is reason to fear, however, that the proportion of breakdowns may be prohibitively high, in which case the mechanism cannot be used in its present state of development and the estimation of its performance parameters is of little value. Moreover, because the experiment will be costly in time, personnel, and equipment, considerable saving will result from discontinuing the experiment at an early stage if the mechanism exhibits too high a failure rate. An experiment is therefore desired that incorporates the following provisions: (1) If the failure rate exceeds a certain tolerance standard, there must be a preassigned degree of assurance of rejecting the mechanism and terminating the experiment; (2) if the failure rate is below a certain acceptable standard, there must be a preassigned degree of assurance of completing the experiment; (3) the experiment should cease at the earliest point at which the accumulated evidence is sufficient to reject the mechanism with the required degree of assurance.

355

## 1.2. Given Features of the Experiment

Previous investigation of the mechanism appears to justify the assumption that the performance characteristic under study is normally distributed about some unknown average value. Moreover, from provisional figures it appears that a satisfactory balance between the cost of the experiment and the precision of the estimates can be obtained by a $2 \times 2 \times 2 \times 2$ factorial experiment involving duplicate observations, that is, from 32 observations. Certain details of the experiment can thus be specified at once. If the mechanism has an acceptable failure rate, no fewer than 32 items can be tested, since 32 measurements of performance are required; on the other hand, no more than 32 *functioning* items need be examined, since these will yield the required number of measurements for estimation with the minimum amount of testing. Thus the criterion for accepting the mechanism can be represented graphically as a straight line by plotting the number of items that fail against the number of items tested. This acceptance line intersects the base line at 32, showing that no fewer than 32 items can be tested; it has a slope of 1, showing that if failures occur, the number of items tested shall be only sufficient to yield 32 performance measurements. As the experiment proceeds, a point can be plotted after each test, representing the cumulated number of failures and the number of tests made; when such a point reaches the acceptance line the experiment is complete, for 32 performance measurements have been obtained.

## 1.3. Statistical Problem

If several hundred trials are made without obtaining the 32 performance measurements needed to complete the experiment, it is obvious that the failure rate is excessive and the experiment ought to be abandoned. This suggests the advisability of introducing a rejection line, showing for each number of items tested the number of failures that would justify rejecting the mechanism and abandoning the experiment; the rejection line should intersect the vertical axis above 0 and have a slope less than 1. The two lines together yield a closed region such that, if the point representing the number of failures and the number of items tested crosses the upper line, the mechanism is rejected and the experiment abandoned; while if the point crosses the lower line, the mechanism is accepted and estimates are made from the resulting 32 measurements.

Since the lower line is fixed by the conditions of the problem, the statistical problem that remains is clearly to determine the upper line in such a way as to meet preassigned requirements on the probabilities of abandonment and completion, with a minimum average amount of testing. No general solution to this problem has been found. The following

discussion will show, however, that sequential principles* can be used to give a good approximation to the solution.

To formulate the problem more precisely, we assume that a proportion of failures $p_2$ exists that is considered unacceptable; that is, if the failure rate equals or exceeds $p_2$, we wish to reject the mechanism. If the failure rate is less than $p_2$, however, we are willing to accept the mechanism; and in particular, if it is less than some value $p_1$, we wish a high degree of assurance of accepting the mechanism.
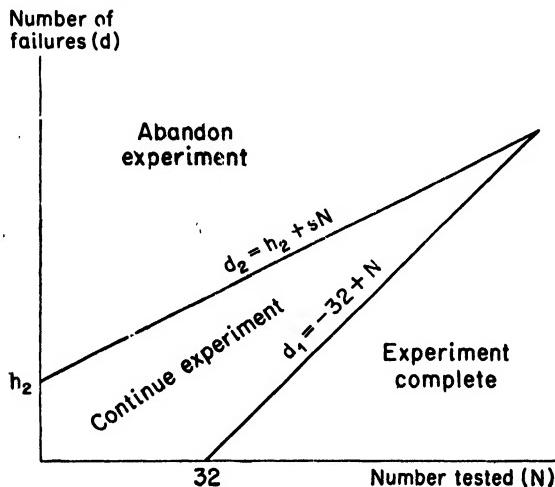


FIG. 12.1.

## 2. PROPOSED SOLUTION

### 2.1. Approximating the Rejection Line by Sequential Principles

The use of sequential principles to meet these requirements is suggested by the desire to minimize the necessary amount of sampling. A sequential criterion makes it possible to reappraise the failure rate of the mechanism after each test and to determine on the basis of the accumulated evidence whether (a) to reject the mechanism, (b) to accept it, or (c) to continue the experiment. As soon as the accumulated evidence is sufficient for decision (a) or (b), the experiment stops and further, unnecessary experimenting is avoided. The usual sequential test, however, consists of both an acceptance criterion and a rejection criterion, whereas completion of the test under consideration requires only a rejection criterion. What we seek, therefore, is a sequential test which makes no provision for acceptance but which achieves certain assigned probabilities of rejection for the specified failure rates $p_1$ and $p_2$. Having found such a test, we can provide for actual acceptance by combining

* See Chap. 17 for a brief account of sequential analysis.

our sequential rejection criterion with the predetermined acceptance criterion, taking care to allow for the effect of this combination on the total probabilities of rejection.

The desired sequential rejection criterion can be derived as a special case of the ordinary sequential binomial test. The ordinary test can be represented on a chart similar to Fig. 12.1 by a pair of parallel straight lines having a slope $s$ and intercepts $-h_1$ and $h_2$. A sequential test which provides only for rejection, that is, which consists solely of the upper decision line, may be regarded as a limiting case of an ordinary sequential binomial test with $h_1 = \infty$. This conception enables us to use the formulas for the usual two-decision test to obtain the operating-characteristic (OC) curve of a one-decision sequential test, that is, the probability of acceptance $L_p$ as a function of the true population proportion of failures $p$. When $h_1 = \infty$, the OC curve is given by*

$$p = \frac{R_p^{-\frac{s}{h_2}} - 1}{R_p^{-\frac{1}{h_2}} - 1} \qquad \text{when } p \le s \qquad (1)$$

and by

$$R_p = 1 - L_p = 1 \qquad \text{when } p > s \qquad (2)$$

Thus, the probability of rejecting is 1 if $p > s$; it lies between 0 and 1 if $p \le s$. Accordingly, we take the value $p_2$, representing the failure rate above which we want to abandon the experiment, as the slope $s$ of the one-decision sequential test. Then if $\alpha$ is the risk we are willing to run of abandoning the experiment and rejecting the mechanism when the true failure rate is only $p_1$, we may substitute $p_1$ for $p$, $\alpha$ for $R_p$, and $p_2$ for $s$ in equation (1) and solve for $h_2$.†

### 2.2. Adjusting the Rejection Line to Take Account of the Acceptance Line

Having determined the one-decision sequential test that yields the requisite probabilities of rejection, we now take into account the acceptance criterion resulting from the conditions of the experiment. This acceptance criterion affects the probabilities in two respects. (1) The probabilities of rejection will be decreased, since those samples which

---

* See Statistical Research Group, *Sequential Analysis*, Appendix B, equations (B.80) and (B.81). (For full citations, see Sec. 5.)

† An approximation to $h_2$ can be obtained by replacing $p_2$ of the formula

$$h_2 = \frac{\log \dfrac{1}{\alpha}}{\log \dfrac{p_2}{p_1}\left(\dfrac{1 - p_1}{1 - p_2}\right)}$$

(*Sequential Analysis*, Sec. 2, equation 2.204) by $p_2 + (p_2 - p_1) = 2p_2 - p_1$.

would lead to late rejection will not be rejected at all (in fact, will now be accepted); (2) for every value of $p < 1$, whether or not $p > s$, the probability of acceptance will be greater than zero. The net effect of the acceptance line is thus to increase the chances of acceptance for all values of $p < 1$. This means that $L_{p_1} (= 1 - R_{p_1})$ will now be positive; while $R_{p_1}$ will be less than $\alpha$. The decrease in $R_{p_1}$ is, of course, desirable, but it can be attained only at the cost of additional testing beyond that necessary to maintain $R_{p_1}$ at the required level $\alpha$. It remains therefore to readjust the decision lines so that $R_{p_1}$ will be $\alpha$ and $L_{p_2}$ will be at an acceptably low level, say $\beta$.

Since the acceptance line is fixed by the conditions of the problem, the necessary changes in the decision lines must be made entirely in the rejection line. It is apparent that the probabilities of rejection can be increased either by reducing $h_2$ or by reducing $s$, or both. It is not clear, however, what combination of adjustments will give the desired risks $\alpha$ and $\beta$ with a minimum amount of inspection. We therefore resort to trial and error to find a combination of $h_2$ and $s$ that gives the desired probabilities of acceptance, leaving unsettled the question whether this combination of $h_2$ and $s$ is the best possible from the standpoint of the average amount of testing. Fortunately, this trial-and-error process can be carried out without too much difficulty, for a straightforward method exists for computing the OC curve for any binomial test of the sort here described. Thus, by determining the OC curve corresponding to each of a number of combinations of $h_2$ and $s$, that combination can be chosen which gives the requisite probabilities of acceptance at $p_1$ and $p_2$. This procedure will yield a plan for the experiment which, though probably not optimum from the standpoint of minimizing inspection, will serve as an approximation to the type of plan desired, that is, lead to early abandonment of the experiment if the true failure rate is high; moreover, the OC curve will be available, so the consequences of using the plan will be fully known.

### 3. AN EXAMPLE

To illustrate the method of solution outlined in Sec. 2, let us suppose that if as many as half the mechanisms are failures we wish a high degree of assurance of abandoning the experiment and rejecting the proposed model, if as few as 40 percent are failures we wish a high degree of assurance of completing the experiment. Let the desired probability of completing the experiment if the failure rate is as high as 50 percent be $\beta = 0.10$, and the probability of abandoning the experiment if the failure rate is as low as 40 percent be $\alpha = 0.10$.

The lower, or acceptance line is determined by the requirements for estimating the performance parameters, as explained above; its equation

is

$$d_1 = -32 + N \tag{3}$$

where $d_1$ represents the cumulated number of failures and $N$ represents the number of tests made.

The first approximation to the rejection line is given by the one-decision sequential test for which $s = 0.5$ and $h_2$ is obtained from

$$0.4 = \frac{0.1^{-\frac{0.5}{h_2}} - 1}{0.1^{-\frac{1}{h_2}} - 1} \tag{4}$$

Writing

$$x = 0.1^{-\frac{1}{h_2}} \tag{5}$$

we have

$$0.4 = \frac{x^{0.5} - 1}{x - 1} = \frac{1}{x^{0.5} + 1} \tag{6}$$

$$x^{0.5} = 1.5 \tag{7}$$

$$x = 2.25 \tag{8}$$

Then, from (5),

$$h_2 = \frac{-\log 0.1}{\log x} \tag{9}$$

$$= \frac{1 \cdot}{\log 2.25}$$

$$= 2.84$$

A moment's reflection on the orientation of the line $d_2 = 2.84 + 0.5N$ on the graph relating number of failures to number of items tested shows that so long as $s = 0.5$, all values of $h$ satisfying the inequality $2.5 < h \leq 3$ are equivalent, since observed values of $d$ and $N$ must be integers. In view of this, and because we expect this first approximation to yield too large probabilities of acceptance, we place the initial value of $h_2$ at 2.5. The first approximation to our test is thus defined by the two straight lines with equations

$$d_2 = 2.5 + 0.5N \quad \text{(rejection line)} \tag{10}$$
$$d_1 = -32 + N \quad \text{(acceptance line)} \tag{11}$$

However, this combination fails to give a satisfactory OC curve, the values of $L_p$ being excessive throughout the range of $p$. On the other hand, the combination $h_2 = 1.6$ and $s = 0.4$ gives values of $L_p$ that are too small, and so also do the combinations $h_2 = 1.5$, $s = 0.5$ and $h_2 = 2.2$, $s = 0.4$. A fifth set, $h_2 = 2.5$ and $s = 0.4$, gives the following OC curve, which, though departing from that originally specified, appears satisfactory.

| Proportion of failures | Probability of completing experiment |
|:---:|:---:|
| 0.1 | 1.000 − |
| 0.2 | 0.990 |
| 0.3 | 0.904 |
| 0.4 | 0.532 |
| 0.5 | 0.102 |
| 0.6 | 0.003 |

## 4. CALCULATION OF OC CURVE

The method of calculating the OC curve for any particular experimental plan of the type described is discussed at length elsewhere,* so it will be outlined here only in general terms.

The basis of the method is the fact that in any truncated sequential test, the number of possible outcomes in sampling is finite and, in practical cases, fairly small. It is therefore possible to enumerate all possible outcomes of the test and to determine the total probability of those outcomes which lead to acceptance, that is, the value of $L_p$.

This calculation is accomplished step by step, as follows: For the first stage of the experiment, all possible ways in which the sample might turn out are specified, and the probability of each outcome is computed on the assumption that the true failure rate of the mechanism is some particular value, say $p_1$. These outcomes are then classified into two groups, those which lead to acceptance and those which lead to the decision to continue testing. The sum of the probabilities of the first class gives the total probability of acceptance at the first stage of the experiment. Each outcome in the second class is then considered as a possible point of departure for the next stage of the experiment; and for each such point of departure, the possible outcomes of the second stage are considered and classified as before. The sum, over all points of departure, of the probabilities of those outcomes which lead to acceptance at the second stage then gives the total probability of acceptance at stage two, while the sum of such probabilities over both stages gives the total probability of acceptance for the first two stages. This basic method of calculation can be carried through the remaining (finite number of) stages of the experiment.

The total probability of acceptance taken over all stages represents the probability $L_{p_1}$ that the experiment will lead to acceptance when the true failure rate is $p_1$. By systematic methods these calculations can be duplicated for a choice of different values of $p$, so that a sequence of values of $p$ and $L_p$ sufficient to determine the entire OC curve, or at least those points of principal interest on the OC curve, can be determined. It is in this fashion that the points on the OC curve in Sec. 3 were determined.

---

* Statistical Research Group, *Sampling Inspection*, Chap. 17, Sec. 2, pp. 187–202.

This method of attack yields only one of many possible tests of the sort described, all having the same probabilities $L_{p_1} = 0.90$ and $L_{p_2} = 0.10$. The optimum test may thus be viewed as the particular one of these tests which requires the smallest average amount of inspection for the relevant range of $p$.  By sufficient calculation this test might be determined by trial and error, since the probabilities used to determine the OC curve of a test suffice to calculate the curve representing the average amount of sampling as a function of the failure rate $p$ (the ASN curve).* If, therefore, we were to determine a whole class of tests with the same $L_{p_1}$ and $L_{p_2}$, we could choose that one whose ASN curve showed the least average cost in sampling in the relevant range of $p$.  However, the volume of calculation necessary to use this approach renders it impractical.

## 5. REFERENCES

Statistical Research Group, Columbia University: *Sampling Inspection*, McGraw-Hill Book Company, Inc., New York, 1948.
———: *Sequential Analysis of Statistical Data: Applications*, rev. ed., Columbia University Press, New York, 1945.

* For a method of calculating the ASN curve in such cases, see Statistical Research Group, *Sampling Inspection*, Chap. 17, Sec. 3.2, pp. 203–209.

# CHAPTER 13

# PLANNING EXPERIMENTS SEEKING MAXIMA

This chapter was prepared by
Milton Friedman, University of Chicago
L. J. Savage, University of Chicago

# CHAPTER 13

## PLANNING EXPERIMENTS SEEKING MAXIMA

### 1. TYPE OF PROBLEM

This chapter deals almost entirely with a special class of experiments, namely, those designed to determine the conditions for which some function attains a maximum. For example, experiments may be conducted on the breaking strength of a type of alloy to determine the composition producing maximum strength. It is not the primary object of such experiments to determine the functional relation between strength and composition, nor even to measure the strength, maximum or otherwise; the primary object is to find a particular combination of the independent variables at which the maximum of the dependent variable is nearly attained. Although the planning of experiments has become a major branch of statistics, little attention has been given to this important class of problem, which is especially common in engineering and development work.*

For concreteness, this chapter considers the specific problem of planning a metallurgical experiment to find the best composition of a certain type of alloy for maximizing the time to rupture under a given stress at a given elevated temperature. Clearly, the principles elucidated in terms of this particular problem are applicable to a wide class of problems.

### 2. PURPOSES OF EXPERIMENT

The problem considered is one in which a single observable result (number of hours it takes a specimen to rupture) is a function of many independent variables (method of treating the alloy, chemical composition of the alloy, and so forth). An experiment is to be conducted in which the result is to be measured for a series of values of the independent variables.

Such an experiment is likely to have two somewhat different purposes: (1) To find the values of the independent variables that make the function (that is, the observable result) a maximum; (2) to evaluate the function in the neighborhood of the maximum. At the present stage of metallurgical knowledge, finding the maximum is usually the primary purpose. There is reason to believe that nothing like the maximum

---

* The pathbreaking work on the problem is that of Hotelling, "Experimental Determination of the Maximum of a Function." (For full citations, see Sec. 6.)

possible strength has yet been attained; and the theory of the subject is so undeveloped as to make direct experiment the most fruitful approach toward finding stronger alloys.  Moreover, there is relatively little interest in knowing the behavior of the function in regions rather distant from the maximum.

If the function has more than one maximum, the ultimate objective is to find the largest maximum—the absolute maximum.  In general, however, this objective is unattainable without complete knowledge of the entire function.  Such knowledge can seldom be secured experimentally, and almost certainly not in a single experiment.  The discussion that follows therefore assumes a more modest objective, the determination of a local maximum.

Although the location of a maximum is of primary importance, mapping the function in the neighborhood of a maximum is important for several reasons:

(1) Typically, maximization of the particular dependent variable measured in a given experiment is only one of several desiderata in choosing an alloy, so it is important to know how much would be sacrificed if the independent variables were altered to satisfy other requirements.  For example, the combination of independent variables that gives the highest time to rupture may give insufficient ductility.  It might then be desirable to know how much strength would be sacrificed by various changes designed to increase ductility.

(2) Knowledge of the relative importance of different independent variables is valuable in designing other experiments to find other maxima.

(3) Knowledge of the function in the region of a maximum may contribute to an understanding of the mechanism at work and may thus advance theoretical knowledge.

(4) It may not be possible to control the independent variables as closely in production as in laboratory experiments.  Knowledge of the function in the neighborhood of a maximum will make it possible to judge the seriousness of departures from control.

### 3. DEFICIENCIES OF FACTORIAL DESIGN

The factorial design is ideally suited for experiments whose purpose is to map a function in a preassigned range of values of the independent variables.*  The fact that the region is preassigned means that it is

---

* The term "factorial design" is used in this chapter to refer to an experimental design in which a limited number of levels of each independent variable are selected in advance and at least one test is made for each combination of the selected levels of the independent variables.  See Fisher, *Design of Experiments*, Chap. VI, The Factorial Design in Experimentation.

The factorial design is in a sense the extreme example of a class of designs having in common the selection in advance of the values of the independent variables at which

feasible to lay out the design completely and specifically in advance of the experiment; the fact that the behavior of the function is desired for all possible combinations of the independent variables means that the experiment must represent all combinations in some way, which a factorial design does.

For the purposes described in the preceding section, however, the factorial design has certain deficiencies:

(1) It devotes observations to exploring regions that may turn out, in the light of the results, to be of no interest because they are far from a maximum.

(2) Because it seeks to explore a given region comprehensively, it must either explore only a small region, or explore a large region very superficially. If a small region is explored, it may turn out to contain no local maximum. The experiment would then at best indicate the direction in which to move in conducting a new experiment. If a large region is explored, it may be explored so superficially that a maximum may be missed entirely, or, if found, will be known so imperfectly that an additional experiment is required.

(3) The factorial design has the disadvantage, somewhat less fundamental than the preceding, that it neglects the fact that some of the variables are continuous and some, though discrete, are ordered. This deficiency could, however, be remedied while retaining the basic principles of the factorial design.

### 4. A SEQUENTIAL DESIGN

These deficiencies of the factorial design suggest that an efficient design for the present purpose ought to be sequential; that is, ought to adjust the experimental program at each stage in the light of the results of prior stages. This section outlines a particular sequential design, primarily to illustrate the direction that seems desirable in the design of metallurgical experiments of the type here considered.

*Step* 1. Select some particular combination of independent variables as a starting point. The combination selected should be the best advance estimate of the optimum combination.

---

observations are to be made. Other designs in this class may, in any particular instance, be more economical than a complete factorial and preferable to it. Economy can be gained without loss, however, only if additional knowledge is available in advance (for example, knowledge that certain interactions are negligible).

Many of the remarks in this chapter are applicable to these other designs as well as to the complete factorial. However, the chapter considers explicitly only the complete factorial design in order to have a sharp and definite contrast and to avoid complications and qualifications that have little bearing on the main issue raised in this chapter.

*Step* 2.   Arrange the independent variables in some order.

*Step* 3.   Make tests at the original starting point and at a series of points differing only in the value assigned the first independent variable, the other variables being held constant.*   From this series of experiments, estimate the value of the first independent variable that maximizes time to rupture.†   If the independent variable is continuous, it will probably be best to find the maximum by some such device as fitting a polynomial to the observations.

*Step* 4.   Holding the first independent variable fixed at the maximizing value found in step 3 and the other independent variables, except the second, at the values originally assigned to them, repeat step 3 with the second independent variable.   Repeat this process with the third, fourth, . . . , independent variable.

*Step* 5.   After all variables have been varied (call this a "round" of experiments), the whole process can be repeated, except that the original starting point is replaced by the combination of values of the independent variables reached at the end of the first round.

Successive rounds will clearly converge to the neighborhood of a local maximum, with perhaps a few "pathological" exceptions; for, aside from statistical fluctuations, each round and each step within a round necessarily represent an improvement.   The rate at which the improvements decline from round to round is likely to give some idea of how much further progress is possible and some indication of the value of additional rounds.   The process of convergence may perhaps be speeded up after one or more complete rounds.   For example, after the second complete

---

* For the present purpose the precise meaning to be attached to "the other variables being held constant" is irrelevant.   For variables representing chemical composition, it can be interpreted as meaning that the relative proportions of the other variables are held constant.   Alternatively, for all but one variable, it can be interpreted as meaning that the percentage a particular element forms of the total is held constant.   Any combination of these two will also do, provided the method of holding other variables constant is followed consistently.

† If the independent variable is ordered, this process also could be sequentialized, with obvious advantages.   For example, tests could be made (*a*) at the original starting point and (*b*) at a single other value of the first independent variable.   If the tests at (*b*) gave a (significantly) higher time to rupture than the tests at (*a*), other tests would be made by moving still farther in the same direction from the original starting point.   If the tests at (*b*) gave a shorter time to rupture, the next tests would be made by moving in the opposite direction from the starting point.   This process would be continued until a value was found higher than its neighbors in both directions.   This aspect of a sequential procedure is not considered in detail here because it is not germane to our main point and because practical considerations having to do with full use of experimental equipment may make it undesirable to sequentialize the experiment fully.   Even though this phase of the experiment is not sequentialized, there is a real problem about designing it efficiently.   See HOTELLING, "Experimental Determination of the Maximum of a Function."

round it might be desirable to move along the vector defined by the end points of the first two rounds, instead of varying the variables according to the original pattern.    After one such step, the original pattern could be used again.

It will clearly be desirable in the early rounds to use rather broad intervals in varying each independent variable and to locate the maximum only very approximately.    Finer and finer intervals should be used in later rounds as the maximum is approached.    A consequence of this is that the process will provide only a very rough mapping of the function in regions far removed from a maximum, but a relatively fine mapping in regions near a maximum.    The mapping of the region in the neighborhood of a maximum can, if desired, be filled in by a supplementary experiment, perhaps of factorial design, after the neighborhood of the maximum has been delimited with sufficient accuracy.

It may be worth repeating that nothing more is claimed for this outline of a sequential design than that it is suggestive.    Numerous problems would arise in applying it to an actual experiment, and there are many possibilities of increasing its efficiency.    The suggested sequential design is in large part simply an abstract and systematic statement of a rather natural, obvious, and common method of locating a maximum or minimum of a function of many variables.

### 5. COMPARISON OF FACTORIAL AND SEQUENTIAL DESIGNS

#### 5.1. Number of Observations

The great advantage of a sequential design over a factorial design is that it will concentrate observations in the neighborhood of a maximum. This advantage is likely to be greater, the larger the number of independent variables and the larger the number of values of each independent variable at which observations are taken.    Let $n$ be the number of variables and suppose that observations are to be taken at $k$ values of each variable (for the sequential design, these need not be the same values in successive rounds).    Then a complete factorial design without replication will require $k^n$ observations.    A single round of the sequential design without replication will require $kn$ observations.    The observations required for a factorial design will therefore *permit $k^n/kn = k^{n-1}/n$* rounds, which increases with $n$ and with $k$.    The number of rounds *required* under the sequential design will decrease as $k$ increases, since the maximum at each step would be determined more accurately.    We do not know how the number of rounds required under the sequential design varies with $n$—nor indeed is it easy to see how to state this question precisely.    We expect, however, that the number of rounds required will not depend very sharply on $n$ but will depend primarily on (a) the close-

ness of the original starting point to the optimum combination and (b) the size of the experimental error relative to the differences it is considered important to detect.

Consider a simple $5 \times 3 \times 3 \times 3$ factorial design, that is, one involving four independent variables, with five values of the first and three values of each of the others. Such a design requires 135 observations, without providing for replications. A single round of the sequential design, covering the same number of values of each independent variable, would require 14 observations. It would therefore be possible to make almost 10 complete rounds with the number of observations required for the factorial experiment. Fewer than 10 rounds should suffice if the original starting point is reasonably close to the maximum. More rounds might be required if the original starting point were far from a maximum; but if so little information about the location of a maximum were available, a factorial experiment might well prove fruitless.

As another example, consider an experiment to determine the optimum composition of the high-carbon vitallium-type alloys with respect to elements that have been included in various tests of this type of alloys. At least ten elements have been included in one or another trial composition. This gives nine independent variables, since the sum of the percentages must add up to 100. Suppose we are willing to use only three levels of each element. A complete factorial design without replications would then require $3^9$, or 19,683, observations. A single round of the sequential design without replications would require only 27 observations, so that it would be possible to make 729 rounds with the number of observations required for a complete factorial design. Five rounds, with duplicate observations at each level on each round, would require only 270 observations, or less than 1.5 per cent of the number required for a complete factorial design without replications. As this example indicates, the advantage of the sequential design in terms of number of observations is likely to become extremely great as the number of variables increases. And, it should be noted, the metallurgical problem in question does in fact involve many variables. The example cited is certainly conservative, since, in addition to the ten chemical elements, the strength of the alloy is likely to be affected by the solution and aging treatments, the methods of casting or of forging, and so forth.

It should be recognized that the additional observations obtained in the factorial experiment are not useless. They provide information about the behavior of the function in regions removed from a maximum, about the relative importance of the variables, and about their interactions. This information is likely to be valuable in choosing new regions within which to search for maxima and in deciding on the vari-

ables on which to concentrate in preliminary experimentation. More important still, the observations some distance from a maximum may, under some circumstances, tell something about the behavior of time to rupture in the neighborhood of that maximum; perhaps the most common circumstances when this occurs are when (*a*) interactions are small, or (*b*) there is available a fairly good approximation to the mathematical form of the function, involving relatively few parameters.

With respect to proviso (*a*), work on metallurgical problems suggests that interactions rarely are small. Furthermore, while the efficiency of a factorial design for finding a maximum is greater if there are no interactions than if there are interactions, so is that of a sequential design. If there are no interactions, a single round of the sequential experiment will, except for statistical fluctuations, lead directly to the maximum. Statistical fluctuations mean that successive rounds will not lead to exactly the same point; but they will tend to give much the same time to rupture, and it should become clear fairly quickly that one is in the immediate neighborhood of a maximum. Moreover, in the absence of interactions, data from successive rounds can be combined to give a better estimate of the maximum.

Proviso (*b*), that there be available a good approximation, involving relatively few parameters, to the mathematical form of the function, is rarely satisfied. But if it is, neither the standard factorial design nor the sequential design suggested in Sec. 4 is appropriate. The problem then is to determine as accurately as possible the values of the independent variables that maximize this function. These maximizing values depend upon the parameters of the function; the optimum design for this purpose thus depends on the form of the function and on the approximate values of its parameters. Probably the best plan is to conduct a small preliminary experiment to get a general idea of the values of the parameters and to use this information in designing a larger experiment.*

### 5.2. Time Trends

One real advantage of the factorial design is that it is possible, through proper randomization, to prevent the introduction of bias from any time trends that may be present. Time trends will, indeed, reduce the power of the factorial experiment to detect the effects that it is designed to measure, but they will not bias the conclusions. The sequential plan, on the other hand, is very much at the mercy of time trends; and great care would have to be taken to see that none crept in. Measures that might be adopted would include (*a*) running control specimens of constant composition along with the test specimens; (*b*) keeping a running record

* See HOTELLING, "Experimental Determination of the Maximum of a Function."

of differences between duplicate specimens to detect changes in experimental error.

## 6. REFERENCES

FISHER, R. A.: *The Design of Experiments*, 3d ed., Oliver and Boyd, Ltd., Edinburgh and London, 1942.

HOTELLING, HAROLD: "The Experimental Determination of the Maximum of a Function," *Annals of Mathematical Statistics*, Vol. 12 (1941), pp. 20–45.

# *Part* III

## TECHNIQUES AND TABLES

# CHAPTER 14

# PROBABILITY THAT SAMPLE MEANS ARE IN OPPOSITE ORDER TO POPULATION MEANS

## FIGURE

## TABLE

This chapter was prepared by
Churchill Eisenhart, National Bureau of Standards

# CHAPTER 14

# PROBABILITY THAT SAMPLE MEANS ARE IN OPPOSITE ORDER TO POPULATION MEANS

## 1. INTRODUCTION

When $N$ objects are drawn independently at ·andom from each of two populations in which some measurable characteristic $x$ of the objects is distributed about mean values $m_1$ and $m_2$, what is the probability that the larger of the two sample means $\bar{x}_1$ and $\bar{x}_2$ will correspond with the population having the smaller mean? In other words, if a decision about the sign of $m_1 - m_4$ is based on the sign of $\bar{x}_1 - \bar{x}_2$, what is the probability that the decision will be wrong?

Such a question arises if two alternative sources $S_1$ and $S_2$ (for example, processes for making, procedures for treating, genetic strains) yield objects of a certain kind, and the worth of each object is determined by the value of some measurable characteristic $x$ that varies about an average $m_1$ for objects from $S_1$ and about an average $m_2$ for objects from $S_2$. If $m_1$ and $m_2$ were known, a decision to adopt $S_1$ or $S_2$ would be determined (completely) by the sign of $m_1 - m_2$; that is, $S_1$ or $S_2$ would be chosen according as $m_1$ exceeds or does not exceed $m_2$. When $m_1$ and $m_2$ are unknown, it is natural to base the decision on the sign of $\bar{x}_1 - \bar{x}_2$, where $\bar{x}_1$ and $\bar{x}_2$ are the average values of $x$ for $N$ objects drawn independently at random from $S_1$ and $S_2$, respectively. Such decisions will be wrong, that is, opposite to the decision that would be made if $m_1$ and $m_2$ were known, whenever the sign of $\bar{x}_1 - \bar{x}_2$ is opposite to that of $m_1 - m_2$. Consequently, it is of interest to know the probability of such an occurrence and how this probability depends on the sample size $N$ and on the absolute magnitude of the true difference, $|m_1 - m_2|$.

In Sec. 2 the probability of $\bar{x}_1 - \bar{x}_2$ having its sign opposite to that of $m_1 - m_2$ is evaluated, and its dependence on $N$ and $|m_1 - m_2|$ studied, for the case where $x$ is normally distributed about $m_1$ with a standard deviation of $\sigma_1$ for objects from $S_1$, or about $m_2$ with standard deviation $\sigma_2$ for objects from $S_2$. In Sec. 3 attention is given to the question of how large $N$ should be in order that the risk of a wrong decision be less than $\beta$ whenever $|m_2 - m_1|$ exceeds $D$, say. The results given in Secs. 2 and 3 are applicable, to a good approximation, in cases where $x$ is non-normally distributed for objects from $S_1$ or $S_2$, provided the distributions of $x$ have finite standard deviations, which is generally the case in prac-

tical work, and $N$ is sufficiently large.    For sample sizes of 4 or larger the approximation will generally be adequate for most practical purposes when $\sigma_1$ and $\sigma_2$ are known.    When $\sigma_1$ and $\sigma_2$ are unknown and $N$ is not less than 30, $\sigma_1$ and $\sigma_2$ may be replaced by their estimates from the samples

$$
s_i = \sqrt{\frac{N \sum_{j=1}^{N} x_{ij}^2 - \left(\sum_{j=1}^{N} x_{ij}\right)^2}{N(N-1)}} \qquad (i = 1, 2) \tag{1}
$$

and the results of Secs. 2 and 3 will not be seriously affected.    In Sec. 4, the conditions under which the sign of $\bar{x}_1 - \bar{x}_2$ provides the "best" decision function for choosing between $S_1$ and $S_2$ are stated, and the problem of choosing a best decision function in other circumstances is discussed briefly.

## 2. PROBABILITY FUNCTION

Let $\bar{x}_1$ and $\bar{x}_2$ denote the means of $N$ observations drawn independently at random from each of two normal populations with means $m_1$ and $m_2$ and standard deviations $\sigma_1$ and $\sigma_2$, respectively.    Then in repeated sampling $\bar{x}_1 - \bar{x}_2$ will be normally distributed about a mean $m_1 - m_2$

TABLE 14.1

PROBABILITY $P(\delta, N)$ THAT SAMPLE MEANS WILL BE IN OPPOSITE ORDER TO POPULATION MEANS FOR SELECTED VALUES OF $\delta$ AND $N$

| Absolute difference between population means in $\sigma$ units* ($\delta$) | Number of observations in each sample ($N$) | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 2 | 4 | 6 | 8 | 10 | 12 | 16 | 20 | 50 | 100 | 200 |
| 0.00 | 0.500 | 0.500 | 0.500 | 0.500 | 0.500 | 0.500 | 0.500 | 0.500 | 0.500 | 0.500 | 0.500 |
| 0.05 | 0.480 | 0.472 | 0.465 | 0.460 | 0.455 | 0.451 | 0.444 | 0.437 | 0.401 | 0.362 | 0.309 |
| 0.10 | 0.460 | 0.444 | 0.431 | 0.421 | 0.412 | 0.403 | 0.389 | 0.376 | 0.309 | 0.240 | 0.159 |
| 0.15 | 0.440 | 0.416 | 0.398 | 0.382 | 0.369 | 0.357 | 0.336 | 0.318 | 0.227 | 0.144 | 0.067 |
| 0 20 | 0.421 | 0.389 | 0.365 | 0.345 | 0.327 | 0.312 | 0.286 | 0.264 | 0.159 | 0.079 | 0.023 |
| 0.25 | 0.401 | 0.362 | 0.333 | 0.309 | 0.288 | 0.270 | 0.240 | 0.215 | 0.106 | 0.039 | 0.006 |
| 0.30 | 0.382 | 0.336 | 0.302 | 0.274 | 0.251 | 0.231 | 0.198 | 0.171 | 0.067 | 0.017 | 0.001 |
| 0.40 | 0.345 | 0.286 | 0.244 | 0.212 | 0.186 | 0.164 | 0.129 | 0.103 | 0.023 | 0.002 | 0.0000 |
| 0.50 | 0.309 | 0.240 | 0.193 | 0.159 | 0.132 | 0.110 | 0.079 | 0.057 | 0.006 | 0.0002 | |
| 0.75 | 0.227 | 0.144 | 0.097 | 0.067 | 0.047 | 0.033 | 0.017 | 0.009 | 0.0001 | 0.0000 | |
| 1.00 | 0.159 | 0.079 | 0.042 | 0.023 | 0.013 | 0.007 | 0.002 | 0.001 | 0.0000 | | |
| 1.25 | 0.106 | 0.039 | 0.015 | 0.006 | 0.003 | 0.001 | 0.0002 | 0.0000 | | | |
| 1.50 | 0.067 | 0.017 | 0.005 | 0.001 | 0.0004 | 0.0001 | 0.0000 | | | | |
| 1.75 | 0.040 | 0.007 | 0.001 | 0.0002 | 0.0000 | 0.0000 | | | | | |
| 2.00 | 0.023 | 0.002 | 0.0003 | 0.0000 | | | | | | | |
| 2.50 | 0.006 | 0.0002 | 0.0000 | | | | | | | | |
| 3.00 | 0.001 | 0.0000 | | | | | | | | | |

* Let $m_1$ and $\sigma_1$ denote the mean and standard deviation of the first population, $m_2$ and $\sigma_2$ denote the mean and standard deviation of the second population, and let $\sigma^2 = (\sigma_1^2 + \sigma_2^2)/2$; then $\delta = |m_1 - m_2|/\sigma$.

with standard deviation $\sigma \sqrt{2/N}$ where

$$\sigma^2 = \frac{\sigma_1^2 + \sigma_2^2}{2} \tag{2}$$

Consequently, the probability that the sign of $\bar{x}_1 - \bar{x}_2$ will be opposite to that of $m_1 - m_2$ is

$$P(\delta, N) = \frac{1}{\sqrt{2\pi}} \int_a^\infty e^{-\frac{t^2}{2}} dt \tag{3}$$

where

$$\delta = \frac{|m_1 - m_2|}{\sigma} \tag{4}$$

and

$$a = \delta \sqrt{\frac{N}{2}} \tag{5}$$

$P(\delta, N)$ is tabulated to 3 decimals (occasionally 4) in Table 14.1 for $\delta = 0$ to 0.30 by steps of 0.05, 0.30 to 0.50 by steps of 0.10, 0.50 to 2.00 by steps of 0.25, 2.00 to 3.00 by steps of 0.50; and for $N = 2$ to 12 by
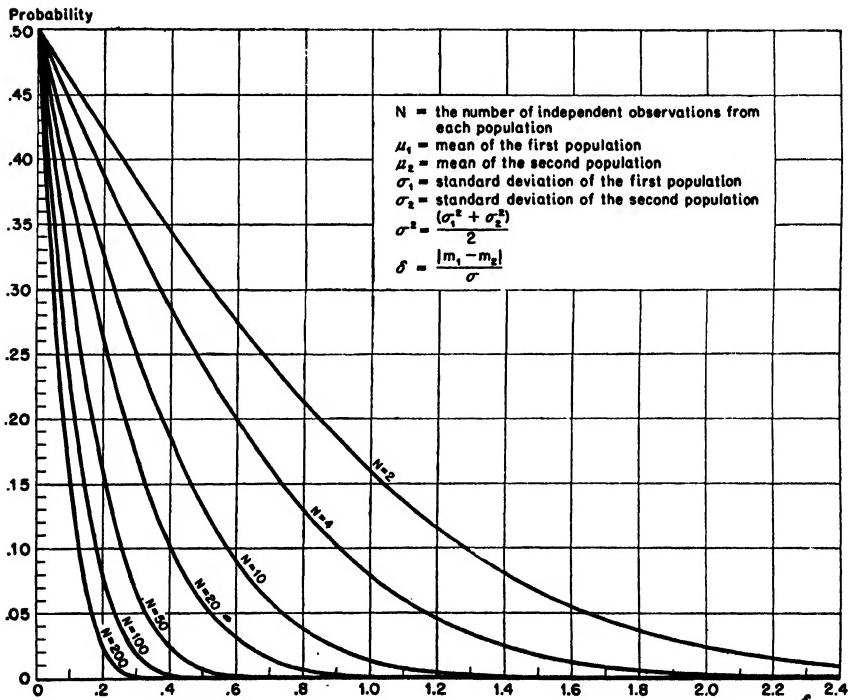


FIG. 14.1.—Probability that means of samples from two normal populations will be in opposite order to population means.

steps of 2, 16, 20, 50, 100, 200.   For example, $P(0.50,10) = 0.132$, which
signifies that, with samples of 10 observations from each of two popula-
tions whose means differ by $\sigma/2$, the probability of the sample means
being in opposite order to the population means is 0.132.   This event
may be expected to occur, therefore, more than 1 time out of 8 in the
long run, under such circumstances.   Clearly it is risky to base decisions
on the sign of $\bar{x}_1 - \bar{x}_2$ for samples of only 10 observations each when
differences of $\sigma/2$, or less, between population means are of practical
importance.

In order to portray better the dependence of $P(\delta,N)$ on $\delta$ and on $N$,
$P(\delta,N)$ has been plotted against $\delta$ in Fig. 14.1 for $N = 2, 4, 10, 20, 50$,
100, and 200.   From these curves it is clear that decisions based on the
sign of $\bar{x}_1 - \bar{x}_2$ will rarely be incorrect when $|m_1 - m_2|$ exceeds $2\sigma$ ($\delta = 2$),
even for samples of only 4 observations each.   On the other hand, if an
incorrect decision is of practical importance when $|m_1 - m_2| = \sigma/2$, and
if it is desired that the probability of an incorrect decision not exceed
0.01 under these circumstances, then about 50 observations from each
population will be needed as a basis for decisions.

### 3. SAMPLE SIZES

If we choose a probability $\beta$ of making an incorrect decision (a
decision based on the sign of $\bar{x}_1 - \bar{x}_2$ different from that which would be
made if the sign of $m_1 - m_2$ were known) and define $K_\epsilon$ by

$$\frac{1}{\sqrt{2\pi}} \int_{K_\epsilon}^{\infty} e^{-\frac{t^2}{2}} dt = \epsilon \tag{6}$$

then, from (3), (4), and (5) it follows that

$$\frac{|m_1 - m_2|}{\sigma} \sqrt{\frac{N}{2}} = K_\beta \tag{7}$$

determines, for a given $N$ and $\sigma$, the value of $|m_1 - m_2|$ for which the prob-
ability of an incorrect decision is $\beta$; or, for a given value of $|m_1 - m_2|/\sigma$,
the value of $N$ that will be needed if the probability of an incorrect
decision is to be $\beta$.   Values of $K_\beta$ can be obtained from Table 1.1 of
Chap. 1, or values of $\beta$ corresponding with various values of $K$ from
Table 1.5.

For example, if an incorrect decision is serious when $|m_1 - m_2| = \sigma/2$,
and if it is desired to limit the risk of such an error to 0.01 under these
circumstances, then (7) implies that at least 44 ($8 \times 2.3263^2 = 43.3$)
observations will be needed from each of the populations.   On the other
hand, if the sampling has already been carried out, taking only 30
observations from each population, it may be some consolation to

know that the risk of an incorrect decision is only 0.05 when

$$|m_1 - m_2| = 1.6449\sigma/\sqrt{15} = 0.42\sigma$$

and is 0.01 when

$$|m_1 - m_2| = 2.3263\sigma/\sqrt{15} = 0.60\sigma.$$

### 4. CONDITIONS UNDER WHICH THE SIGN OF $\bar{x}_1 - \bar{x}_2$ PROVIDES THE BEST DECISION FUNCTION

In Sec. 1 it was stated that if a decision to choose $S_1$ or $S_2$ would be determined (completely) by the sign of $m_1 - m_2$ if this sign were known, it is natural to base the decision on the sign of $\bar{x}_1 - \bar{x}_2$. This raises the question of whether the sign of $\bar{x}_1 - \bar{x}_2$ is actually the best function of the data upon which to base the decision

H. A. Simon has shown that when (a) $x$ is normally distributed in each of the populations $S_1$ and $S_2$, (b) the standard deviation of $x$ is the same ($\sigma_1 = \sigma_2$) in $S_1$ and $S_2$; (c) there is no a priori basis for a choice between $S_1$ and $S_2$ without knowledge of the values of $m_1$ and $m_2$ (or, at least, of the sign of $m_1 - m_2$); and (d) the choice between $S_1$ and $S_2$ is to be based on some symmetric function of the observed values of an equal number $N$ of observations drawn independently at random from $S_1$ and $S_2$, then the probability of an incorrect decision will be a minimum, whatever the values of $m_1$ and $m_2$ and of the common value of $\sigma_1$ and $\sigma_2$, if the decision to choose $S_1$ or $S_2$ is based on the sign of $\bar{x}_1 - \bar{x}_2$. As we have seen above, the probability of an incorrect decision, when the sign of $\bar{x}_1 - \bar{x}_2$ provides the basis for decision, is determined completely by $N$, $\sigma = \sigma_1 = \sigma_2$, and $|m_1 - m_2|$ under these conditions.

When conditions (a), (c), and (d) above are fulfilled, but $\sigma_1 \neq \sigma_2$, then, Simon has found, there does not exist any symmetric function of the $2N$ observations having the property that when the decision to choose between $S_1$ and $S_2$ is based on the observed value of this function the probability of an incorrect decision will be a minimum whatever the values of $m_1$ and $m_2$; this is true even when the values of $\sigma_1$ and $\sigma_2$ are known. This is the case a fortiori when the value of either $\sigma_1$ or $\sigma_2$ is unknown. He has found, however, that if it is agreed that (d') the choice between $S_1$ and $S_2$ is to be based on some symmetric function of the differences $x_{11} - x_{21}$, $x_{12} - x_{22}$, $\cdots$, $x_{1N} - x_{2N}$, then the sign of $\bar{x}_1 - \bar{x}_2$ is the decision function that minimizes the probability of error, whatever the values of $m_1$ and $m_2$ and of $\sigma_1$ and $\sigma_2$.

When $x$ is non-normally distributed in one or both of the populations, then $\bar{x}_1 - \bar{x}_2$ no longer furnishes the most efficient estimate of $m_1 - m_2$ and the sign of $\bar{x}_1 - \bar{x}_2$ will not be the best decision function upon which to base a choice between $S_1$ and $S_2$, even when $\sigma_1 = \sigma_2$. Welch has discussed the determination of the best discriminant function upon which

to base a choice between two populations when the form of the distribution of $x$ in each of the populations is known.

## 5. REFERENCES

SIMON, HERBERT A.: "Symmetric Tests of the Hypothesis That the Mean of One Normal Population Exceeds That of Another," *Annals of Mathematical Statistics*, Vol. 14 (1943), pp. 149–154.

WELCH, B. L.: "Note on Discriminant Functions," *Biometrika*, Vol. 31 (1939), pp. 218–220.

# CHAPTER 15

## SIGNIFICANCE OF THE LARGEST OF A SET OF SAMPLE ESTIMATES OF VARIANCE

This chapter was prepared by
Churchill Eisenhart, National Bureau of Standards
Herbert Solomon, Office of Naval Research

## CHAPTER 15

## SIGNIFICANCE OF THE LARGEST OF A SET OF SAMPLE ESTIMATES OF VARIANCE

### 1. TYPE OF PROBLEM

When data relate to several different processes, to several different products, to a single process at different times, or to a particular type of product from each of several different sources, the analyst is frequently faced with the necessity of deciding whether the observed differences in variability of the data within the respective classes are consistent with the supposition that the intrinsic variability is uniform, that is, the same for each class.

If the classes are not uniform in variability, then simple comparisons of the classes in other respects may be misleading. Thus, if the within-class variability differs greatly from class to class, the mean values for the respective classes of the variable under consideration no longer provide an adequate basis for comparison, since a small mean associated with a large dispersion may imply greater frequency of occurrence of undesirable values of the characteristic than will a larger mean associated with a smaller dispersion of individual values. In such cases it may be more useful to compare the classes with respect to the probability of occurrence of an individual value greater than a specified upper limit $U$ (as in Chap. 1) or with respect to some other feature that takes into account both the average value and the variability of the characteristic in each class. In this connection it should be recalled that the customary $F$ test of the homogeneity of class means, which consists of comparing the observed ratio of the between-class mean square to the pooled within-class mean square, is valid only when the intrinsic within-class variability is the same for each class. Indeed, a pooled estimate of within-class variability, obtained by combining measures of the respective within-class dispersions, may be useless for practical purposes, unless there is homogeneity of within-class variability or the method of combining the within-class dispersions takes proper account of the relative importance (for example, the relative frequency of occurrence) of the respective classes under realistic conditions.

Of the various criteria that have been advanced for testing the homogeneity of a set of variance estimates, the test proposed by Bartlett is generally considered to be the best one available of the "omnibus"

variety.* An omnibus test is one that has good discriminating power with respect to a large variety of alternatives to the null hypothesis. The customary $F$ test of the homogeneity of a set of means is often used as an omnibus test. This is the case when there are $k$ means and the between-means mean square having $k - 1$ degrees of freedom is compared with the error mean square. But "the statistician who does not know in advance with which type of alternative to [homogeneity] he may be faced is in the position of a carpenter who is summoned to a house to undertake a job of an unknown kind and is only able to take one tool with him! Which shall it be? Even if there is an 'omnibus' tool, it is likely to be far less sensitive at any particular job than a specialized one; but the specialized tool will be quite useless under the wrong conditions."† It is therefore advisable in applications of the $F$ test to isolate and test individually the mean squares corresponding to comparisons that have special interest a priori, or that represent comparisons between rational subgroups of the data that may have unequal (population) means for various a priori reasons. By thus adapting the $F$ test to the circumstances of the study involved, its discriminating power is improved, at least with respect to the alternatives to homogeneity that are believed to be the most likely to exist or to be of special importance if they do exist. Bartlett has pointed out that his over-all test for homogeneity of a set of variances can be separated, in similar fashion, into parts representing comparisons of special interest between subsets of variance estimates, and the statistical significance of each part tested individually.‡ Like the $F$ test for homogeneity of means, Bartlett's test has been shown to be unbiased and to possess various other desirable properties.§ When the issue is homogeneity of variances, Bartlett's test assumes the role played by the $F$ test when the issue is homogeneity of means.

Bartlett's test, though applicable, is somewhat inappropriate when the alternative to homogeneity that is most likely to represent the true situation, or that is of special importance when it does obtain, is that the expected value of some one of the mean squares (variance estimates) substantially exceeds the expected values of the others, which may be homogeneous or not. By "some one" it is meant that prior to the

---

* BARTLETT, "Sufficiency and Statistical Tests," especially p. 273. Also see BISHOP and NAIR, "Homogeneity of a Set of Estimated Variances"; PITMAN, "Location and Scale Parameters"; and HARTLEY, "Homogeneity of a Set of Variances." Bartlett's test is described in the next section of this chapter; also in Rider, *Statistical Methods*, Chap. 7, Sec. 51, where an illustration is given. (For full citations, see Sec. 5.)

† PEARSON, "Efficiency of Statistical Tools," p. 316.

‡ BARTLETT, "Examples of Statistical Methods," especially pp. 158–159.

§ PITMAN, "Location and Scale Parameters," especially Secs. 2 to 6; also HARTLEY, "Homogeneity of a Set of Variances."

analysis of the data—or, better, prior to the collection of the data—there is no basis for identifying the anomalous one of the set. The several variance estimates may correspond, for example, to the several machines of a production line and represent the within-machine variability of products produced by the respective machines, which variability may be expected to be the same from machine to machine unless one of the machines is out of adjustment, in which event the variability of products from this machine will exceed that of products from other machines. Similarly, the several variance estimates may correspond to samples taken from a production line at various times during a shift and may be expected to be homogeneous if the process is in control but to have a high value if the process was out of control on one occasion. In situations of this type what is needed is a test of the statistical significance of the amount by which the largest of the variance estimates exceeds the others. Cochran has developed such a test and has provided a small table of 5 percent significance levels to facilitate its use.*

Cochran's test for the statistical significance of the largest of a set of variance estimates is described and its use illustrated in Sec. 3. Table 15.1, which gives 5 percent levels of the test, is an extension of Cochran's table. Table 15.2 gives 1 percent levels, which were not tabulated by Cochran. In view of the approximations used in their computation, many of the entries in these tables are not accurate to 4 decimal places, as tabulated. Their accuracy is believed to be sufficient for most practical purposes, at least until more accurate values are available. For discussion of the methods of computation see Sec. 4.

### 2. BARTLETT'S TEST

Let $s_1^2, s_2^2, \ldots, s_k^2$ be $k$ variance estimates distributed independently as $\chi^2 \sigma^2 / n$ for $n_1, n_2, \ldots, n_k$ degrees of freedom, respectively, and $s^2$ be the pooled variance of the $k$ samples

$$s^2 = \frac{\Sigma n_i s_i^2}{n} \tag{1}$$

where

$$n = \Sigma n_i \tag{2}$$

Bartlett's test involves computing

$$B = \frac{1}{C}\left( n \log_e s^2 - \sum n_i \log_e s_i^2 \right) \tag{3}$$

$$= \frac{2.30259}{C}\left( n \log_{10} s^2 - \sum n_i \log_{10} s_i^2 \right)$$

---

* Cochran, "Largest of a Set of Estimated Variances."

**where**

$$C = 1 + \frac{\sum \frac{1}{n_i} - \frac{1}{n}}{3(k-1)} \cdot \qquad (4)$$

Then even for quite small values of the $n_i$, say, 5 or more, $B$ has approximately the $\chi^2$ distribution with $k-1$ degrees of freedom. Since $C > 1$, it need not be calculated if for $C = 1$ the value of $B$ is not significant.*

### 3. COCHRAN'S TEST

### 3.1. Tables

Let $s_1^2$, $s_2^2$, . . . , $s_k^2$ be $k$ variance estimates distributed independently as $\chi^2\sigma^2/n$ for $n$ degrees of freedom each, and let

$$g = \frac{\text{largest } s^2}{s_1^2 + s_2^2 + \cdots + s_k^2} \qquad (5)$$

be the ratio of the largest $s^2$ to their total. Table 15.1 gives values of $g$ to 4 decimals that are exceeded with probability $P$, where $0.04875 < P \leq 0.05000$, for $n = 1$ to 10, 16, 36, 144, $\infty$, and $k = 2$ to 10, 12, 15, 20, 24, 30, 40, 60, 120, $\infty$. Table 15.2 gives values of $g$ to 4 decimals that are exceeded with probability $P$, where $0.00995 < P \leq 0.01000$, for the same values of $k$ and $n$. These two tables thus show the 5 percent and 1 percent levels of $g$. They are exact levels (to 4-decimal accuracy) whenever they exceed 0.5000 and in many other instances also. For example, a calculation of Cochran's implies that the entries in Table 15.1 for $n \leq 10$ and $k \leq 10$ are exact 5 percent levels to 4-place accuracy. Presumably the entries of Table 15.2 for $n \leq 10$ and $k \leq 10$ are likewise exact 1 percent levels to 4-place accuracy. Values of $n$ and $k$ exceeding 10 were selected for accuracy in interpolation. Thus $n$-wise interpolation should be carried out in terms of $12/\sqrt{n}$; $k$-wise interpolation, in terms of $120/k$ or $60/k$.

### 3.2. Example

The following are the squares of sample standard deviations obtained from 20 sets of six strength tests of cement mortar, each set being obtained from a different mixing of the same cement mortar under similar condi-

---

* Since this chapter went to press, Thompson and Merrington have provided "Tables for Testing the Homogeneity of a Set of Estimated Variances." Writing (3) above as $B = M/C$, they have tabled maximum and minimum values of the true 5 percent and 1 percent significance levels of $M$ for $k = 3$ to 15 and for

$$c_1 = 3(k-1)(C-1)$$

from 0 to 5 by steps of 0.5; 5 to 10 by steps of 1; 12; and 14.

tions.* Are these 20 squared standard deviations homogeneous? In

| Sample Number | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| (S.D.)² | 271 | 389 | 617 | 107 | 1052 | 340 | 383 | 508 | 511 | 438 |
| Sample Number | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| (S.D.)² | 280 | 209 | 936 | 281 | 346 | 383 | 492 | 512 | 462 | 349 |

particular, is the standard deviation computed from the fifth sample "out of line"?

In this case $n = 5$, $k = 20$, and $g = 1,052/8,866 = 0.1187$, which is less than the corresponding 5 percent level (0.1735) given in Table 15.1. The anomalous value of the standard deviation from the fifth sample may, therefore, be attributed to chance, and the 20 sets of strength tests may be regarded as homogeneous with respect to variability of the strength determinations.

The squared standard deviations given above correspond to the definition

$$\text{(Sample standard deviation)}^2 = \frac{\sum_{i=1}^{6} (x_i - \bar{x})^2}{6} \tag{6}$$

where $\bar{x} = \text{sample average} = \frac{1}{6} \sum_{i=1}^{6} x_i$, the $x_i$ ($i = 1, 2, \cdots, 6$) being the 6 individual strength determinations of the sample concerned, in pounds per square inch. Consequently, each value is $5/6$ of the corresponding variance estimate $s^2$, defined by

$$s^2 = \frac{\sum_{i=1}^{6} (x_i - \bar{x})^2}{5} \tag{7}$$

However, the factor $5/6$ is common to the numerator and denominator of $g$, so need not be taken into account explicitly.

### 4. COMPUTATION OF TABLES

Cochran has shown† that if $s_1^2$, $s_2^2$, . . . , $s_k^2$ are variance estimates distributed independently as $\chi^2 \sigma^2/n$ for $n$ degrees of freedom, and if

$$g_i = \frac{s_i^2}{s_1^2 + s_2^2 + \cdots + s_k^2} \qquad (i = 1, 2, \cdots, k) \tag{8}$$

* DAVIES and PEARSON, "Estimating from Samples," Table IV, p. 92; a description of the tests appears on p. 77.

† Cochran's $r$ and $n$ correspond to our $n$ and $k$, respectively.

TABLE 15.1

UPPER 5 PERCENT LEVELS OF THE RATIO OF THE LARGEST TO THE SUM OF $k$ INDEPENDENT ESTIMATES OF VARIANCE, EACH OF WHICH IS BASED ON $n$ DEGREES OF FREEDOM

| $k$ \ $n$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 16 | 36 | 144 | ∞ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 0.9985 | 0.9750 | 0.9392 | 0.9057 | 0.8772 | 0.8534 | 0.8332 | 0.8159 | 0.8010 | 0.7880 | 0.7341 | 0.6602 | 0.5813 | 0.5000 |
| 3 | 0.9669 | 0.8709 | 0.7977 | 0.7457 | 0.7071 | 0.6771 | 0.6530 | 0.6333 | 0.6167 | 0.6025 | 0.5466 | 0.4748 | 0.4031 | 0.3333 |
| 4 | 0.9065 | 0.7679 | 0.6841 | 0.6287 | 0.5895 | 0.5598 | 0.5365 | 0.5175 | 0.5017 | 0.4884 | 0.4366 | 0.3720 | 0.3093 | 0.2500 |
| 5 | 0.8412 | 0.6838 | 0.5981 | 0.5441 | 0.5065 | 0.4783 | 0.4564 | 0.4387 | 0.4241 | 0.4118 | 0.3645 | 0.3066 | 0.2513 | 0.2000 |
| 6 | 0.7808 | 0.6161 | 0.5321 | 0.4803 | 0.4447 | 0.4184 | 0.3980 | 0.3817 | 0.3682 | 0.3568 | 0.3135 | 0.2612 | 0.2119 | 0.1667 |
| 7 | 0.7271 | 0.5612 | 0.4800 | 0.4307 | 0.3974 | 0.3726 | 0.3535 | 0.3384 | 0.3259 | 0.3154 | 0.2756 | 0.2278 | 0.1833 | 0.1429 |
| 8 | 0.6798 | 0.5157 | 0.4377 | 0.3910 | 0.3595 | 0.3362 | 0.3185 | 0.3043 | 0.2926 | 0.2829 | 0.2462 | 0.2022 | 0.1616 | 0.1250 |
| 9 | 0.6385 | 0.4775 | 0.4027 | 0.3584 | 0.3286 | 0.3067 | 0.2901 | 0.2768 | 0.2659 | 0.2568 | 0.2226 | 0.1820 | 0.1446 | 0.1111 |
| 10 | 0.6020 | 0.4450 | 0.3733 | 0.3311 | 0.3029 | 0.2823 | 0.2666 | 0.2541 | 0.2439 | 0.2353 | 0.2032 | 0.1655 | 0.1308 | 0.1000 |
| 12 | 0.5410 | 0.3924 | 0.3264 | 0.2880 | 0.2624 | 0.2439 | 0.2299 | 0.2187 | 0.2098 | 0.2020 | 0.1737 | 0.1403 | 0.1100 | 0.0833 |
| 15 | 0.4709 | 0.3346 | 0.2758 | 0.2419 | 0.2195 | 0.2034 | 0.1911 | 0.1815 | 0.1736 | 0.1671 | 0.1429 | 0.1144 | 0.0889 | 0.0667 |
| 20 | 0.3894 | 0.2705 | 0.2205 | 0.1921 | 0.1735 | 0.1602 | 0.1501 | 0.1422 | 0.1357 | 0.1303 | 0.1108 | 0.0879 | 0.0675 | 0.0500 |
| 24 | 0.3434 | 0.2354 | 0.1907 | 0.1656 | 0.1493 | 0.1374 | 0.1286 | 0.1216 | 0.1160 | 0.1113 | 0.0942 | 0.0743 | 0.0567 | 0.0417 |
| 30 | 0.2929 | 0.1980 | 0.1593 | 0.1377 | 0.1237 | 0.1137 | 0.1061 | 0.1002 | 0.0958 | 0.0921 | 0.0771 | 0.0604 | 0.0457 | 0.0333 |
| 40 | 0.2370 | 0.1576 | 0.1259 | 0.1082 | 0.0968 | 0.0887 | 0.0827 | 0.0780 | 0.0745 | 0.0713 | 0.0595 | 0.0462 | 0.0347 | 0.0250 |
| 60 | 0.1737 | 0.1131 | 0.0895 | 0.0765 | 0.0682 | 0.0623 | 0.0583 | 0.0552 | 0.0520 | 0.0497 | 0.0411 | 0.0316 | 0.0234 | 0.0167 |
| 120 | 0.0998 | 0.0632 | 0.0495 | 0.0419 | 0.0371 | 0.0337 | 0.0312 | 0.0292 | 0.0279 | 0.0266 | 0.0218 | 0.0165 | 0.0120 | 0.0083 |
| ∞ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Let $s_1^2, s_2^2, \ldots, s_k^2$ be any set of $k$ quantities each independently distributed as $\chi^2 \sigma^2 / n$ for $n$ degrees of freedom, and let

$$g = \frac{(\text{largest } s^2)}{\sum_{i=1}^{k} s_i^2}.$$

Then, for a given pair of values of $k$ and $n$, the probability $P$ that $g$ exceeds the corresponding tabular entry satisfies the inequality $0.04875 < P \leq 0.05000$. For further details regarding accuracy of entries, see Sec. 3 of this chapter.

TABLE 15.2    SIGNIFICANCE OF LARGEST VARIANCE    391

### TABLE 15.2

UPPER 1 PERCENT LEVELS OF THE RATIO OF THE LARGEST TO THE SUM OF $k$ INDEPENDENT ESTIMATES OF VARIANCE, EACH OF WHICH IS BASED ON $n$ DEGREES OF FREEDOM

| $k$ \ $n$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 16 | 36 | 144 | ∞ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 0.9999 | 0.9950 | 0.9794 | 0.9586 | 0.9373 | 0.9172 | 0.8988 | 0.8823 | 0.8674 | 0.8539 | 0.7949 | 0.7067 | 0.6062 | 0.5000 |
| 3 | 0.9933 | 0.9423 | 0.8831 | 0.8335 | 0.7933 | 0.7606 | 0.7335 | 0.7107 | 0.6912 | 0.5743 | 0.6059 | 0.5153 | 0.4230 | 0.3333 |
| 4 | 0.9676 | 0.8643 | 0.7814 | 0.7212 | 0.6761 | 0.6410 | 0.6129 | 0.5897 | 0.5702 | 0.5536 | 0.4884 | 0.4057 | 0.3251 | 0.2500 |
| 5 | 0.9279 | 0.7885 | 0.6957 | 0.6329 | 0.5875 | 0.5531 | 0.5259 | 0.5037 | 0.4854 | 0.4697 | 0.4094 | 0.3351 | 0.2644 | 0.2000 |
| 6 | 0.8828 | 0.7218 | 0.6258 | 0.5635 | 0.5195 | 0.4866 | 0.4608 | 0.4401 | 0.4229 | 0.4084 | 0.3529 | 0.2858 | 0.2229 | 0.1667 |
| 7 | 0.8376 | 0.6644 | 0.5685 | 0.5080 | 0.4659 | 0.4347 | 0.4105 | 0.3911 | 0.3751 | 0.3616 | 0.3105 | 0.2494 | 0.1929 | 0.1429 |
| 8 | 0.7945 | 0.6152 | 0.5209 | 0.4627 | 0.4226 | 0.3932 | 0.3704 | 0.3522 | 0.3373 | 0.3248 | 0.2779 | 0.2214 | 0.1700 | 0.1250 |
| 9 | 0.7544 | 0.5727 | 0.4810 | 0.4251 | 0.3870 | 0.3592 | 0.3378 | 0.3207 | 0.3067 | 0.2950 | 0.2514 | 0.1992 | 0.1521 | 0.1111 |
| 10 | 0.7175 | 0.5358 | 0.4469 | 0.3934 | 0.3572 | 0.3308 | 0.3106 | 0.2945 | 0.2813 | 0.2704 | 0.2297 | 0.1811 | 0.1376 | 0.1000 |
| 12 | 0.6528 | 0.4751 | 0.3919 | 0.3428 | 0.3099 | 0.2861 | 0.2680 | 0.2535 | 0.2419 | 0.2320 | 0.1961 | 0.1535 | 0.1157 | 0.0833 |
| 15 | 0.5747 | 0.4069 | 0.3317 | 0.2882 | 0.2593 | 0.2386 | 0.2228 | 0.2104 | 0.2002 | 0.1918 | 0.1612 | 0.1251 | 0.0934 | 0.0667 |
| 20 | 0.4799 | 0.3297 | 0.2654 | 0.2288 | 0.2048 | 0.1877 | 0.1748 | 0.1646 | 0.1567 | 0.1501 | 0.1248 | 0.0960 | 0.0709 | 0.0500 |
| 24 | 0.4247 | 0.2871 | 0.2295 | 0.1970 | 0.1759 | 0.1608 | 0.1495 | 0.1406 | 0.1338 | 0.1283 | 0.1060 | 0.0810 | 0.0595 | 0.0417 |
| 30 | 0.3632 | 0.2412 | 0.1913 | 0.1635 | 0.1454 | 0.1327 | 0.1232 | 0.1157 | 0.1100 | 0.1054 | 0.0867 | 0.0658 | 0.0480 | 0.0333 |
| 40 | 0.2940 | 0.1915 | 0.1508 | 0.1281 | 0.1135 | 0.1033 | 0.0957 | 0.0898 | 0.0853 | 0.0816 | 0.0668 | 0.0503 | 0.0363 | 0.0250 |
| 60 | 0.2151 | 0.1371 | 0.1069 | 0.0902 | 0.0796 | 0.0722 | 0.0668 | 0.0625 | 0.0594 | 0.0567 | 0.0461 | 0.0344 | 0.0245 | 0.0167 |
| 120 | 0.1225 | 0.0759 | 0.0585 | 0.0489 | 0.0429 | 0.0387 | 0.0357 | 0.0334 | 0.0316 | 0.0302 | 0.0242 | 0.0178 | 0.0125 | 0.0083 |
| ∞ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Let $s_1^2, s_2^2, \ldots, s_k^2$ be any set of $k$ quantities each independently distributed as $\chi^2 \sigma^2 / n$ for $n$ degrees of freedom, and let

$$g = \frac{(\text{largest } s^2)}{\sum\limits_{i=1}^{k} s_i^2}$$

Then, for a given pair of values of $k$ and $n$, the probability $P$ that $g$ exceeds the corresponding tabular entry satisfies the inequality $0.00995 < P \leq 0.01000$. For further details regarding accuracy of entries, see Sec. 3 of this chapter.

then the probability that the largest of the ratios (8) exceeds $g$ is given by

$$P(g) = kP_1(g) - \frac{k(k-1)}{2!}P_2(g) + \frac{k(k-1)(k-2)}{3!}P_3(g) - \cdots$$
$$+ (-1)^{k-1}P_k(g) \quad (9)$$

where $P_1(g)$ is the probability that a specified one of the ratios exceeds $g$, $P_2(g)$ the probability that a specified pair of the ratios both exceed $g$, and so on. Clearly

$$P_k(g) = 0 \quad \text{when } g > \frac{1}{k} \quad (10)$$

for it is certain that the largest ratio will be as great as $1/k$; so the number of non-zero terms in (9) is the greatest integer less than $1/g$.

Cochran has shown that

$$P_1(g) = \frac{\int_g^1 x^{\frac{n}{2}-1}(1-x)^{\frac{n(k-1)}{2}-1}\,dx}{B\left[\frac{n}{2}, \frac{n(k-1)}{2}\right]} \quad 0 \le g \le 1 \quad (11)$$

and has given $P_2(g)$ and $P_3(g)$ in multiple integral form, from which the multiple integral form of $P_j(g)$, $j = 4, 5, \cdots, k$, is apparent. He has pointed out that, since $P_2(g) < [P_1(g)]^2$ for all values of $n$ and $k$, the upper $\alpha$ levels of significance of $g$ can be evaluated to a good first approximation using only the first term of (9). Thus, if $\bar{g}_\alpha$ is determined from

$$P_1(\bar{g}_\alpha) = \frac{\alpha}{k} \quad (12)$$

it follows from (9) that

$$\alpha\left(1 - \frac{\alpha}{2}\right) < \alpha - \frac{k-1}{2k}\alpha^2 < P(\bar{g}_\alpha) \le \alpha \quad (13)$$

that is, upper 5 percent levels of $g$ determined from (12) will in all cases correspond to $0.04875 < P \le 0.05000$; and, in view of (10), will be exact when they exceed $\frac{1}{2}$. Similarly, upper 1 percent levels determined from (12) will correspond to $0.00995 < P \le 0.01000$ and will be exact when they exceed $\frac{1}{2}$.

From (11) it follows that

$$\bar{g}_\alpha(n,k) = 1 - x_{\frac{\alpha}{k}} \quad (14)$$

$$= \frac{F_{\frac{\alpha}{k}}}{F_{\frac{\alpha}{k}} + (k-1)}$$

where

$$x_{\frac{\alpha}{k}} = x_{\frac{\alpha}{k}}\left[\frac{n}{2}, \frac{n(k-1)}{2}\right] \quad (15)$$

is the lower $\alpha/k$ level of significance of the beta-function distribution, and

$$F_{\frac{\alpha}{k}} = F_{\frac{\alpha}{k}}[n, n(k-1)] \tag{16}$$

is the upper $\alpha/k$ level of significance of the $F$ distribution. Hence $\bar{g}_{0.05}(n,k)$ can be obtained from the tables of Thompson, or from those of Thompson and Merrington, for $k = 2$, 5, and 10; and $\bar{g}_{0.01}(n,k)$ can be obtained similarly for $k = 2$. The values of $\bar{g}_{0.05}$ given in Table 15.1 for $k = 2$, 5, and 10, and of $\bar{g}_{0.01}$ given in Table 15.2 for $k = 2$, were obtained in this manner.

When $n$ is an even integer, (11) can be reduced to a polynomial in $g$. Thus, when $n = 2$

$$P_1(g) = (1 - g)^{k-1} \tag{17}$$

and when $n = 4$

$$P_1(g) = (1 - g)^{2k-2}[1 + 2(k-1)g] \tag{18}$$

as Cochran has pointed out. The corresponding polynomials for $n = 6$, 8, and 10 were determined. The values of $\bar{g}_{0.05}$ and $\bar{g}_{0.01}$ given in Tables 15.1 and 15.2 for $n = 2$, 4, 6, 8, and 10 were obtained with the aid of these polynomials. In all other cases the values of $\bar{g}_{0.05}$ and $\bar{g}_{0.01}$ were evaluated from $0.05/k$ and $0.01/k$ levels of $F$, which, in turn, were determined by means of a modification of a procedure due to Paulson.

It is believed that as estimates of $\bar{g}_{0.05}$ and $\bar{g}_{0.01}$ the entries of Tables 15.1 and 15.2 are correct to within 1, or possibly 2, units in the fourth decimal place. The accuracy of the tabular entries as estimates of the true 5 percent and 1 percent levels of $g$ is in general unknown, although it is known that the tabular values tend to exceed the true values. It is believed, however, that in all cases the tabular 5 percent levels correspond to $0.04875 < P \leq 0.05000$ and the tabular 1 percent levels to $0.00995 < P \leq 0.01000$. With this understanding, Tables 15.1 and 15.2 should be useful in practical work until more accurate tables become available.

## 5. REFERENCES

BARTLETT, M. S.: "Properties of Sufficiency and Statistical Tests," *Proceedings of the Royal Society of London*, Vol. 160A (1937), pp. 268–282.

———: "Some Examples of Statistical Methods of Research in Agriculture and Applied Biology," *Supplement to the Journal of the Royal Statistical Society*, Vol. 4 (1937), pp. 137–183.

BISHOP, D. J., and NAIR, U. S.: "A Note on Certain Methods of Testing for the Homogeneity of a Set of Estimated Variances," *Supplement to the Journal of the Royal Statistical Society*, Vol. 6 (1939), pp. 89–99.

COCHRAN, W. G.: "The Distribution of the Largest of a Set of Estimated Variances as a Fraction of Their Total," *Annals of Eugenics*, Vol. 11 (1941), pp. 47–52.

DAVIES, O. L., and PEARSON, E. S.: "Methods of Estimating from Samples the Population Standard Deviation," *Supplement to the Journal of the Royal Statistical Society*, Vol. 1 (1934), pp. 76–93.

HARTLEY, H. O.: "Testing the Homogeneity of a Set of Variances," *Biometrika*, Vol. 31 (1939–1940), pp. 249–255.

PAULSON, EDWARD: "An Approximate Normalization of the Analysis of Variance Distribution," *Annals of Mathematical Statistics*, Vol. 13 (1942), pp. 233–235.

PEARSON, E. S., assisted by SEKAR, C. CHANDRA: "The Efficiency of Statistical Tools and a Criterion for the Rejection of Outlying Observations," *Biometrika*, Vol. 28 (1936), pp. 308–320.

PITMAN, E. J. G.: "Tests of Hypotheses concerning Location and Scale Parameters," *Biometrika*, Vol. 31 (1939–1940), pp. 200–215.

RIDER, PAUL R.: *An Introduction to Modern Statistical Methods*, John Wiley & Sons, Inc., New York, 1939, 220 + ix pp.

THOMPSON, CATHERINE M.: "Tables of Percentage Points of the Incomplete Beta Function," *Biometrika*, Vol. 32 (1941), pp. 151–181.

———and MERRINGTON, MAXINE: "Tables of Percentage Points of the Inverted Beta (*F*) Distribution," *Biometrika*, Vol. 33 (1943), pp. 73–88.

———and———: "Tables for Testing the Homogeneity of a Set of Estimated Variances," *Biometrika*, Vol. 33 (1946,) pp. 296–304.

# CHAPTER 16

# INVERSE SINE TRANSFORMATION OF PROPORTIONS

This chapter was prepared by
Churchill Eisenhart, National Bureau of Standards

# CHAPTER 16

# INVERSE SINE TRANSFORMATION OF PROPORTIONS

## 1. HISTORY OF THE TRANSFORMATION

If successive samples of $n$ items each are drawn independently at random from a very large (infinite) population in which a proportion $P$ of the items possess some definite attribute $A$, then the proportions $p_1, p_2, p_3, \ldots$ of items possessing $A$ in the successive samples will be distributed in such a manner that as the number of independent samples is increased without limit the average of the $p$'s will tend to $P$ and the mean of their squared deviations from $P$ will tend to $P(1 - P)/n$. In the language of statistics this is expressed by saying that, if a sample of $n$ items is drawn at random from an infinite population in which a proportion $P$ of the items possess a definite attribute $A$, and if the proportion of items possessing $A$ in the sample is denoted by $p$, then

$$\text{Expected value of } p = E(p) = P \tag{1}$$

and

$$\text{Variance of } p = V(p) = \frac{P(1 - P)}{n} \tag{2}$$

or

$$\text{Standard deviation of } p = \sigma_p = \sqrt{\frac{P(1 - P)}{n}} \tag{3}$$

Equation (3) or equation (2), considered in conjunction with equation (1), states that observed proportions $p$, based on successive independent random samples of size $n$, may be expected to be grouped more closely about the true proportion $P$ when $n$ is large than when $n$ is small; for fixed sample size, their sampling variation about $P$ will be greatest when $P$ equals $\frac{1}{2}$ and will decrease toward zero as either $P$ or $1 - P$ approaches zero. This dependence of the sampling variation of an observed proportion $p$ upon the true proportion that it estimates, $P$, is a source of inconvenience in certain kinds of theoretical work and in some types of statistical analysis of research data. Basically, the inconvenience is a consequence of the fact that for purposes of statistical inference (for example, estimation of parameters, tests of agreement between observation and hypothesis) the worth or "weight" of an individual observation (here, a single observed proportion $p$) is generally inversely proportional to its sampling variance. Accordingly, the for-

mulas of statistical theory and the calculations of statistical analysis are simpler, in general, when the observations involved are of equal weight, that is, of equal precision.

Recognizing this fact, R. A. Fisher, in a paper on theoretical genetics published in 1922,[*] introduced the transformation $\cos \varphi = 1 - 2p$, where $\varphi$ varies from 0 to $\pi$ as $p$ varies from 0 to 1, and utilized the fact that, when $\varphi$ is computed from large samples, the variance of $\varphi$ is, for all practical purposes, independent of the true proportion $P$.

Fisher employed this transformation to simplify the mathematical equations occurring in an application of statistical theory to a problem in genetics. In 1930 he published another paper on this problem[†] in which he considered the bias of the transformation in large samples and showed that his failure to make allowance for this bias in his earlier paper accounted for the discrepancy between one of the results therein and a solution to the same problem obtained in a different manner by Sewall Wright. Not until 1936 were the possibilities of this transformation in connection with statistical analysis of research data discussed in print, when three writers indicated advantages to be gained from its use in statistical analyses of data expressed in the form of observed proportions (or percentages of items counted). Bartlett,[‡] on a suggestion from Yates, Fisher's coworker at the Rothamsted Experimental Station, considered the use of a related transformation ($\sin^2 \theta = p$, where $\theta$ varies from 0 to $\pi/2$; see below) as a means of rendering proportions amenable to analysis-of-variance techniques, which assume equal precision of the observations involved. With the same objective, Bliss,[§] after some correspondence with Fisher, recommended the use of the angular transformation $\sin^2 \theta = p$ in the statistical analysis of field-trial data expressed in percentages (that is, 100 times the proportion of items with a given attribute in a sample) and gave an extensive table of $\theta$ (in degrees) for percentages (that is, $100p$) from 0.01 to 99.99. Zubin[‖] pointed out the advantages of using Fisher's transformation as a means of simplifying the calculations involved in testing the statistical significance of a difference between two observed proportions, his attention having been drawn to Fisher's papers by Hotelling. Zubin, utilizing this transformation, subsequently prepared nomographs[¶] by means of which the statistical significance of the difference between two observed proportions can be determined almost at a glance. These nomographs jointly are, in a sense, the ancestors of the nomograph presented in Chap. 7 above.

---

[*] FISHER, "On the Dominance Ratio," p. 326.   (For full citations, see Sec. 5.1.)
[†] FISHER, "The Distribution of Gene Ratios."
[‡] BARTLETT, "Square Root Transformation in Analysis of Variance," p. 73.
[§] BLISS, "The Analysis of Field Experimental Data Expressed in Percentages."
[‖] ZUBIN, "Note on a Transformation Function for Proportions and Percentages."
[¶] ZUBIN, "Nomographs for Determining the Significance of the Difference."

Since 1936 a great deal has been written on the transformation of proportions to angles, to facilitate or validate various forms of statistical analysis. A fairly complete bibliography is given at the end of this chapter; it includes references to tables that have been published to facilitate the angular transformations in current use.

The remainder of this chapter is devoted to a discussion of the effectiveness of Fisher's transformation in small samples, especially the dependence on $P$ of the expected value and the variance of $\varphi$. A simple adjustment, due to Bartlett,* is discussed, and a table (Table 16.4) of adjusted values provided to facilitate its use. More elaborate adjustments, determined by the data themselves in any given case, have been discussed by Cochran,† and tables‡ are available to facilitate their use. In the last section of this chapter the general method of determining transformations to stabilize variance is outlined, the angular transformation of Fisher derived, and several other such transformations noted. A method of discovering Bartlett's adjustment is noted, which may be useful in other situations.

## 2. NATURE OF THE TRANSFORMATION

Let $p$ denote the proportion of items possessing some particular attribute in a sample of $n$ items drawn independently at random from a very large (infinite) population of items in which a proportion $P$ of the items possess the attribute in question. Under these circumstances the mean, variance, and standard deviation of $p$ are functions of $P$, given by equations (1), (2), and (3). Let $\varphi$ be an angle (in radians) determined by the transformation

$$\varphi = 2 \arcsin \sqrt{p} = \arccos (1 - 2p) \qquad 0 \leq p \leq 1 \qquad 0 \leq \varphi \leq \pi \quad (4)$$

the inverse of which is

$$p = \sin^2 \frac{\varphi}{2} = \frac{1 - \cos \varphi}{2} \qquad 0 \leq p \leq 1 \qquad 0 \leq \varphi \leq \pi \quad (5)$$

Then, when the number of observations $n$ upon which $p$ is based is large,

$$\text{Expected value of } \varphi = E(\varphi) \quad (6)$$
$$= \varphi(P) \text{ approximately}$$

where

$$\varphi(P) = 2 \arcsin \sqrt{P} \quad (7)$$

---

* BARTLETT, "Some Examples of Statistical Methods," p. 168.

† COCHRAN, "The Analysis of Variance When Experimental Errors Follow the Poisson or Binomial Laws," pp. 342 ff.

‡ FISHER and YATES, *Statistical Tables.*

and

$$\text{Variance of } \varphi = V(\varphi) = \frac{1}{n} \text{ approximately*} \qquad (8)$$

An elementary derivation of this transformation is given in Sec. 4. This is the form in which the transformation was introduced by Fisher in 1922; the transformation has been employed in this form by Zubin and, with certain modifications, by the Statistical Research Group.†

Other writers, following Bartlett and Bliss, have transformed $p$ to an angle $\theta$ equal to $\varphi/2$, whence $V(\theta) = 1/4n$ approximately for $\theta$ in radians or $821/n$ approximately for $\theta$ in degrees. Thus, when $\theta$ is expressed in radians, $V(\theta)$ has, for a wide range of values of $P$, approximately the value of $V(p)$ when $P = \frac{1}{2}$. This advantage, if such it is, is lost, however, when $\theta$ is expressed in degrees.

TABLE 16.1

MEANS AND VARIANCES OF VARIOUS FUNCTIONS OF A SAMPLE PROPORTION FOR SELECTED VALUES OF THE POPULATION PROPORTION WHEN THE SAMPLE SIZE IS 10

| $P$ | $\varphi(P)$ | $E(p)$ | $40V(p)$ | $E(\varphi)$ | $10V(\varphi)$ | $E(\varphi_B)$ | $10V(\varphi_B)$ |
|---|---|---|---|---|---|---|---|
| 0.0500 | 0.4510 | 0.0500 | 0.1900 | 0.2856 | 1.2917 | 0.4757 | 0.4481 |
| 0.1000 | 0.6435 | 0.1000 | 0.3600 | 0.5134 | 1.6567 | 0.6241 | 0.7489 |
| 0.1500 | 0.7954 | 0.1500 | 0.5100 | 0.7006 | 1.6529 | 0.7632 | 0.9363 |
| 0.2000 | 0.9273 | 0.2000 | 0.6400 | 0.8598 | 1.5332 | 0.8939 | 1.0436 |
| 0.3000 | 1.1593 | 0.3000 | 0.8400 | 1.1260 | 1.2968 | 1.1350 | 1.1224 |
| 0.5000 | 1.5708 | 0.5000 | 1.0000 | 1.5708 | 1.1433 | 1.5708 | 1.1258 |
| 0.7000 | 1.9823 | 0.7000 | 0.8400 | 2.0156 | 1.2968 | 2.0066 | 1.1224 |
| 0.8000 | 2.2143 | 0.8000 | 0.6400 | 2.2818 | 1.5332 | 2.2477 | 1.0436 |
| 0.8500 | 2.3462 | 0.8500 | 0.5100 | 2.4410 | 1.6529 | 2.3784 | 0.9363 |
| 0.9000 | 2.4981 | 0.9000 | 0.3600 | 2.6282 | 1.6567 | 2.5175 | 0.7489 |
| 0.9500 | 2.6906 | 0.9500 | 0.1900 | 2.8560 | 1.2917 | 2.6659 | 0.4481 |

Explanation

Let     $P$ denote the proportion of items possessing a definite characteristic in an infinite population

$p$ denote the proportion of items possessing this characteristic in a sample of 10 items drawn independently at random from the population

Then $\varphi(p) = 2 \arcsin \sqrt{p} = \arccos(1 - 2p)$, $0 \le p \le 1$, $0 \le \varphi \le \pi$

$\varphi_B(p) = \varphi(p)$ for $0 < p < 1$, $\varphi_B(0) = \varphi(\frac{1}{40})$, $\varphi_B(1) = \pi - \varphi_B(0)$

and $E(\ )$ denotes average or expected value of the quantity within the parentheses

$V(\ )$ denotes variance of the quantity within the parentheses.

* If $\varphi$ is measured in degrees

$$V(\varphi) = \left(\frac{180}{\pi}\right)^2 \frac{1}{n} = \frac{3282.8}{n}$$

† *Sequential Analysis*, Sec. 3.6; also Chap. 7 of this volume.

## 3. EFFECTIVENESS OF THE TRANSFORMATION

Apparently the only published investigation of the effectiveness of the inverse sine transformation in small samples is that of Bartlett* for the case of $n = 10$.   His analysis shows that when $n = 10$, the variance of $\varphi$ depends very markedly on the true proportion $P$. A year later he remarked† that a good correction is to replace $p = 0$ by $p = 1/4n$, and $p = 1$ by $p = 1 - 1/4n$, although he gave no supporting evidence.

Tables 16.1 to 16.3 show, for $n = 10$, 20, and 30, respectively, the values to 4 decimal places of $E(p)$, $4nV(p)$, $\varphi(P)$, $E(\varphi)$, $nV(\varphi)$, $E(\varphi_B)$, and $nV(\varphi_B)$ corresponding to $P = 0.05$ to 0.20 by steps of 0.05, 0.30, 0.50, 0.70, and 0.80 to 0.95 by steps of 0.05, where $\varphi_B$ denotes the angular transformation (4) modified as suggested by Bartlett, that is,

$$\varphi_B(p) = 2 \arcsin \sqrt{p} = \arccos (1 - 2p) \qquad 0 < p < 1 \qquad 0 < \varphi_B < \pi$$
$$\varphi_B(0) = 2 \arcsin \sqrt{1/4n}, \quad \varphi_B(1) = \pi - \varphi_B(0) \tag{9}$$

TABLE 16.2

MEANS AND VARIANCES OF VARIOUS FUNCTIONS OF A SAMPLE PROPORTION FOR SELECTED VALUES OF THE POPULATION PROPORTION WHEN THE SAMPLE SIZE IS 20

| $P$ | $\varphi(P)$ | $E(p)$ | $80V(p)$ | $E(\varphi)$ | $20V(\varphi)$ | $E(\varphi_B)$ | $20V(\varphi_B)$ |
|---|---|---|---|---|---|---|---|
| 0.0500 | 0.4510 | 0.0500 | 0.1900 | 0.3541 | 1.6315 | 0.4344 | 0.7247 |
| 0.1000 | 0.6435 | 0.1000 | 0.3600 | 0.5866 | 1.5458 | 0.6139 | 1.0138 |
| 0.1500 | 0.7954 | 0.1500 | 0.5100 | 0.7602 | 1.3240 | 0.7688 | 1.0974 |
| 0.2000 | 0.9273 | 0.2000 | 0.6400 | 0.9035 | 1.1810 | 0.9061 | 1.1051 |
| 0.3000 | 1.1593 | 0.3000 | 0.8400 | 1.1469 | 1.0858 | 1.1471 | 1.0784 |
| 0.5000 | 1.5708 | 0.5000 | 1.0000 | 1.5708 | 1.0562 | 1.5708 | 1.0561 |
| 0.7000 | 1.9823 | 0.7000 | 0.8400 | 1.9947 | 1.0858 | 1.9945 | 1.0784 |
| 0.8000 | 2.2143 | 0.8000 | 0.6400 | 2.2381 | 1.1810 | 2.2355 | 1.1051 |
| 0.8500 | 2.3462 | 0.8500 | 0.5100 | 2.3814 | 1.3240 | 2.3728 | 1.0974 |
| 0.9000 | 2.4981 | 0.9000 | 0.3600 | 2.5550 | 1.5458 | 2.5277 | 1.0138 |
| 0.9500 | 2.6906 | 0.9500 | 0.1900 | 2.7875 | 1.6315 | 2.7072 | 0.7247 |

Explanation

Let     $P$ denote the proportion of items possessing a definite characteristic in an infinite population

$p$ denote the proportion of items possessing this characteristic in a sample of 20 items drawn independently at random from the population

Then $\varphi(p) = 2 \arcsin \sqrt{p} = \arccos (1 - 2p)$, $0 \leq p \leq 1$, $0 \leq \varphi \leq \pi$

$\varphi_B(p) = \varphi(p)$ for $0 < p < 1$, $\varphi_B(0) = \varphi(\frac{1}{80})$, $\varphi_B(1) = \pi - \varphi_B(0)$

and $E(\ )$ denotes average or expected value of the quantity within the parentheses

$V(\ )$ denotes variance of the quantity within the parentheses.

* BARTLETT, "Square Root Transformation in Analysis of Variance," pp. 74-75.
† BARTLETT, "Some Examples of Statistical Methods," p. 168.

The mean values and variances of $\varphi$ and $\varphi_B$ were determined by direct calculation.*

If the inverse sine transformation completely stabilized variance in small samples, then $nV(\varphi)$ would be unity in all cases, or at least would be constant for a given value of $n$. That this is not the case is clear from Tables 16.1 to 16.3. Evidently, $nV(\varphi_B)$ is somewhat less variable than $nV(\varphi)$ over the range of values of $P$ considered. The nature of the dependence of $nV(\varphi)$ and $nV(\varphi_B)$ on $P$ is depicted graphically in Figs. 16.1 to 16.3, for $n = 10, 20,$ and $30$, respectively. For purposes of comparison $4nV(p)$ has been plotted also. The following conclusions may be drawn:

(1) For $n \geq 10$ and $0.05 \leq P \leq 0.95$, $nV(\varphi)$ and $nV(\varphi_B)$ are less variable than $4nV(p)$.

TABLE 16.3

MEANS AND VARIANCES OF Various FUNCTIONS OF A SAMPLE PROPORTION FOR
SELECTED VALUES OF THE POPULATION PROPORTION WHEN THE SAMPLE SIZE
Is 30

| $P$ | $\varphi(P)$ | $E(p)$ | $120V(p)$ | $E(\varphi)$ | $30V(\varphi)$ | $E(\varphi_B)$ | $30V(\varphi_B)$ |
|---|---|---|---|---|---|---|---|
| 0.0500 | 0.4510 | 0.0500 | 0.1900 | 0.3885 | 1.6443 | 0.4278 | 0.8985 |
| 0.1000 | 0.6435 | 0.1000 | 0.3600 | 0.6119 | 1.3359 | 0.6197 | 1.0921 |
| 0.1500 | 0.7954 | 0.1500 | 0.5100 | 0.7757 | 1.1560 | 0.7771 | 1.0987 |
| 0.2000 | 0.9273 | 0.2000 | 0.6400 | 0.9134 | 1.0885 | 0.9136 | 1.0773 |
| 0.3000 | 1.1593 | 0.3000 | 0.8400 | 1.1516 | 1.0484 | 1.1516 | 1.0481 |
| 0.5000 | 1.5708 | 0.5000 | 1.0000 | 1.5708 | 1.0358 | 1.5708 | 1.0358 |
| 0.7000 | 1.9823 | 0.7000 | 0.8400 | 1.9900 | 1.0484 | 1.9900 | 1.0481 |
| 0.8000 | 2.2143 | 0.8000 | 0.6400 | 2.2282 | 1.0885 | 2.2280 | 1.0773 |
| 0.8500 | 2.3462 | 0.8500 | 0.5100 | 2.3659 | 1.1560 | 2.3645 | 1.0987 |
| 0.9000 | 2.4982 | 0.9000 | 0.3600 | 2.5297 | 1.3359 | 2.5219 | 1.0921 |
| 0.9500 | 2.6806 | 0.9500 | 0.1900 | 2.7531 | 1.6443 | 2.7138 | 0.8985 |

Explanation

Let    $P$ denote the proportion of items possessing a definite characteristic in an infinite population

$p$ denote the proportion of items possessing this characteristic in a sample of 30 items drawn independently at random from the population

Then $\varphi(p) = 2 \arcsin \sqrt{p} = \arccos(1 - 2p)$, $0 \leq p \leq 1$, $0 \leq \varphi \leq \pi$

$\varphi_B(p) = \varphi(p)$ for $0 < p < 1$, $\varphi_B(0) = \varphi(\frac{1}{120})$, $\varphi_B(1) = \pi - \varphi_B(0)$

and $E(\ )$ denotes average or expected value of the quantity within the parentheses

$V(\ )$ denotes variance of the quantity within the parentheses.

* Binomial probabilities to 7 decimals were obtained from *Tables of the Incomplete Beta Function;* arcsin values to 6 decimals were obtained by interpolation in Hayashi's *Sieben-und mehrstellige Tafeln,* the entries of this latter table being corrected in some cases in accordance with the list of corrections given in *Mathematical Tables and Other Aids to Computation,* Vol. 1, No. 6, p. 197.

(2) For $n \geq 20$ and $0.05 \leq P \leq 0.95$, $nV(\varphi_B)$ is less variable than $nV(\varphi)$.

(3) For $n \geq 10$, and $0.075 \leq P \leq 0.925$, $nV(\varphi_B)$ is less variable than $nV(\varphi)$.

Since it is quite apparent from the foregoing that the inverse sine transformation, especially with Bartlett's correction, is advantageous

Let P denote the proportion of items in an infinite population that possess a given characteristic;

  p denote the proportion of items possessing the characteristic among the 10 items drawn independently at random from the population;

  $\Phi(p) = 2\arcsin\sqrt{p} = \arccos(1-2p)$;

  $\Phi_B(p) = \Phi(p)$, $0 < p < 1$; $\Phi_B(0) = \Phi(1/40)$; $\Phi_B(1) = \pi - \Phi_B(0)$;

then  $V_1 = 4P(1-P) = 4 \times 10 \times V(p)$

    $V_2 = 10\, V(\Phi)$

    $V_3 = 10\, V(\Phi_B)$
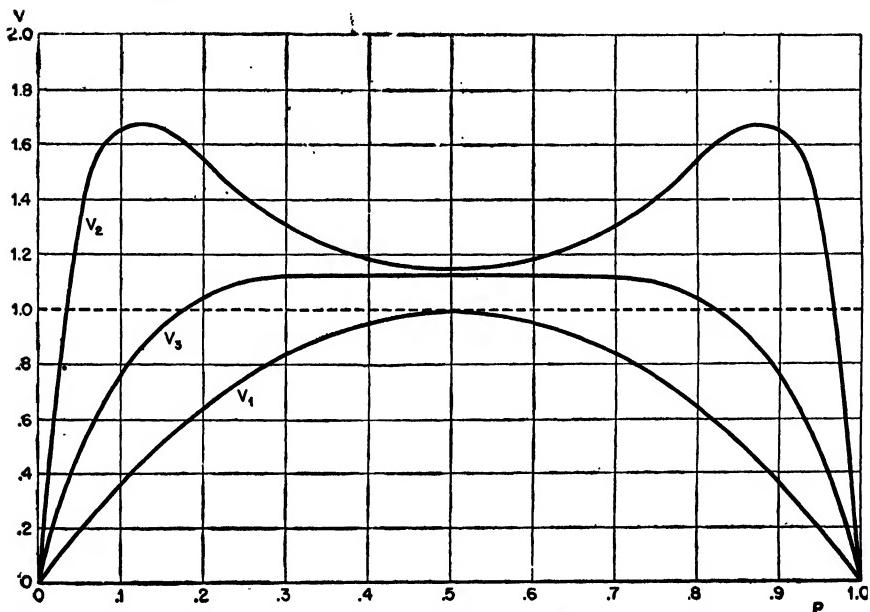


FIG. 16.1.—Dependence of the variances of three functions of a sample proportion $p$ on the population proportion $P$ when the sample size is 10.

from the viewpoint of variance stabilization, its effectiveness in this regard increasing rapidly as the sample size $n$ is increased, the question arises whether estimates of $P$ obtained by transforming back to proportions after averaging in angular units will be biased. A comparison of the values of $\varphi(P)$, $E(\varphi)$, and $E(\varphi_B)$ shown in Tables 16.1 to 16.3 reveals that $E(\varphi)$ and $E(\varphi_B)$ generally are less than $\varphi(P)$ when $P < \frac{1}{2}$;

greater than $\varphi(P)$ when $P > \frac{1}{2}$; and the three are equal when $P = \frac{1}{2}$.* Furthermore, for all values of $n$ and $P$ considered in these tables, $E(\varphi_B)$ is the closer to $\varphi(P)$. From Tables 16.1 to 16.3 it is seen that, when $0.05 \leq P \leq 0.95$, the bias $E(\varphi_B) - \varphi(P)$ is less than 15 per cent of the

Let $P$ denote the proportion of items in an infinite population that possess a given characteristic;

   $p$ denote the proportion of items possessing the characteristic among the 20 items drawn independently at random from the population;

   $\Phi(p) = 2\arcsin \sqrt{p} = \arccos(1-2p)$;

   $\Phi_B(p) = \Phi(p), \ 0 < p < 1; \ \Phi_B(0) = \Phi(1/80); \ \Phi_B(1) = \pi - \Phi_B(0)$;

then    $V_1 = 4P(1-P) = 4 \times 20 \times V(p)$
       $V_2 = 20 \ V(\Phi)$
       $V_3 = 20 \ V(\Phi_B)$
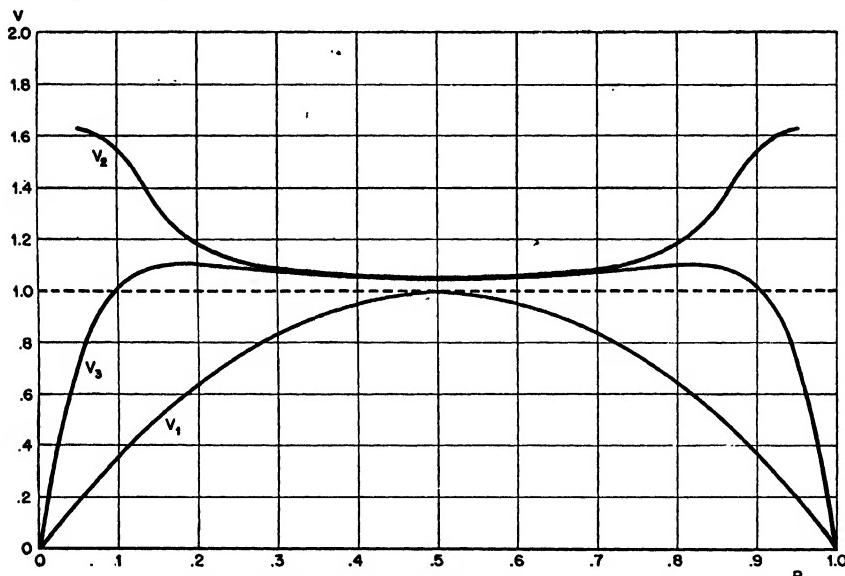


FIG. 16.2.—Dependence of the variances of three functions of a sample proportion $p$ on the population proportion $P$ when the sample size is 20.

standard deviation of $\varphi_B$ when $n \geq 10$ and thus is of negligible magnitude for most practical purposes. Thus Bartlett's correction, in addition to its advantage from the viewpoint of stabilizing variance, tends to reduce to negligible magnitudes the biases of estimates of $P$ obtained by reconverting averages expressed in angular units to estimated mean proportions with the aid of equation (5).

* The values of $E(\varphi_B)$ for $P = 0.05$ and $P = 0.95$ when $n = 10$ are exceptions. Such exceptions can be expected to occur with $\varphi_B$ in those instances in which the $p = 0$ or $p = 1$ class has a large probability of occurrence, that is, when $P$ is close to 0 or 1 and $n$ is small.

Values of $\varphi_B(0)$ and $\varphi_B(1)$, expressed in radians, have been tabulated in Table 16.4 for sample sizes from 10 to 50, in order to facilitate the use of Bartlett's correction.   The values of $\varphi_B(0)$ and $\varphi_B(1)$ given in Table 16.4 may be used in conjunction with Zubin's "Table for Transforming Proportions to $T$ values"* without modification; they should be divided

Let P denote the proportion of items in an infinite population that possess a given characteristic;

p denote the proportion of items possessing the characteristic among the 30 items drawn independently at random from the population;

$\Phi(p) = 2 \arcsin \sqrt{p} = \arccos(1-2p)$;

$\Phi_B(p) = \Phi(p), 0 < p < 1; \Phi_B(0) = \Phi(1/120); \Phi_B(1) = \pi - \Phi_B(0)$;

then  $V_1 = 4 P(1-P) = 4 \times 30 \times V(p)$

$V_2 = 30 V(\Phi)$

$= 30 V(\Phi_B)$
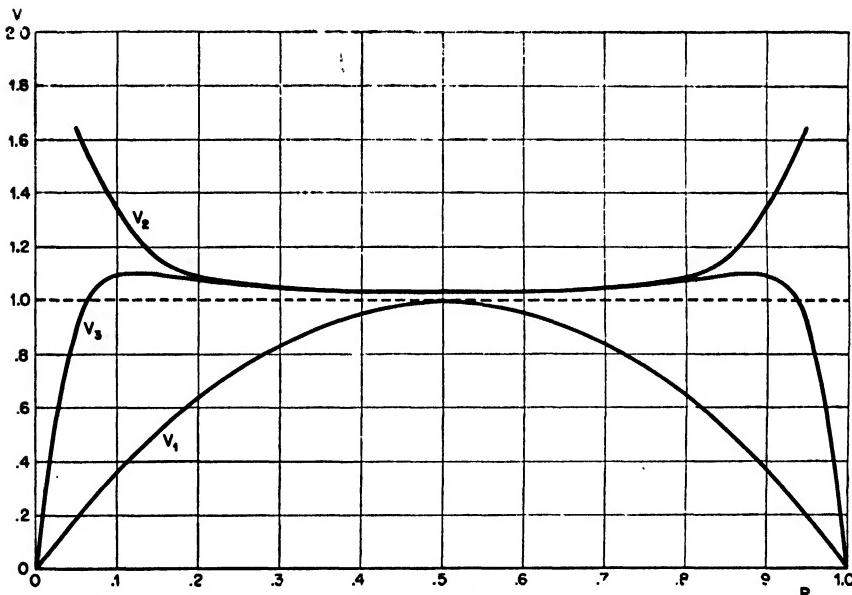


Fig. 16.3.—Dependence of the variances of three functions of a sample proportion $p$ on the population proportion $P$ when the sample size is 30.

by 2 before being used in conjunction with Bartlett's table†; they should be multiplied by 28.64789 before being used in conjunction with the tables of Bliss, Fisher and Yates, and Snedecor.

A previous publication of the Statistical Research Group gives a table‡ of the transformation.

* ZUBIN, "Note on a Transformation," Table 4, p. 220.
† BARTLETT, "Subsampling for Attributes," Table II, p. 132.
‡ Sequential Analysis, Table 3.64, p. 3.57.

$$y(s,n) = 2 \arcsin \sqrt{s/n} \qquad 0 < s < n, 0 < y < \pi$$
$$y(0,n) = \sqrt{2/n} \qquad \text{for } s = 0 \tag{10}$$
$$y(n,n) = \pi - \sqrt{2/n} \qquad \text{for } s = n$$

for values of $n$ from 10 to 50.   Comparison of the foregoing with equation (9) above reveals that $y(s) = \varphi_B(p)$ with $p = s/n$ for $0 < s < n$; $y(0) = \sqrt{2}\varphi_B(0)$ approximately and $y(n) = \pi - \sqrt{2}\varphi_B(0)$ approximately.   Since the publication of this table of the $y$ transformation, it has been found that Bartlett's adjustments, $\varphi_B(0)$ and $\varphi_B(1)$, are slightly more effective for stabilizing variance, especially when $P$ is

TABLE 16.4

VALUES OF $\varphi_B(0)$ AND $\varphi_B(1)$ FOR SAMPLES OF 10 TO 50 ITEMS TOGETHER WITH VALUES OF THE VARIANCE OF $\varphi_B$ WHEN $P = 0.5$

| Sample size, $n$ | $\varphi_B(0)$, radians | $\varphi_B(1)$, radians | $V_{0.5}(\varphi_B)$, radians² | Sample size, $n$ | $\varphi_B(0)$, radians | $\varphi_B(1)$, radians | $V_{0.5}(\varphi_B)$ radians² |
|---|---|---|---|---|---|---|---|
| 10 | 0.3176 | 2.8240 | 0.1126 | 30 | 0.1828 | 2.9588 | 0.0345 |
| 11 | 0.3027 | 2.8389 | 0.1011 | 31 | 0.1798 | 2.9617 | 0.0334 |
| 12 | 0.2897 | 2.8519 | 0.0918 | 32 | 0.1770 | 2.9646 | 0.0323 |
| 13 | 0.2782 | 2.8633 | 0.0840 | 33 | 0.1743 | 2.9673 | 0.0313 |
| 14 | 0.2681 | 2.8735 | 0.0775 | 34 | 0.1717 | 2.9699 | 0.0303 |
| 15 | 0.2589 | 2.8827 | 0.0719 | 35 | 0.1692 | 2.9724 | 0.0294 |
| 16 | 0.2507 | 2.8909 | 0.0670 | 36 | 0.1669 | 2.9747 | 0.0286 |
| 17 | 0.2431 | 2.8985 | 0.0628 | 37 | 0.1646 | 2.9770 | 0.0278 |
| 18 | 0.2363 | 2.9053 | 0.0591 | 38 | 0.1624 | 2.9792 | 0.0270 |
| 19 | 0.2299 | 2.9117 | 0.0558 | 39 | 0.1603 | 2.9813 | 0.0263 |
| 20 | 0.2241 | 2.9175 | 0.0528 | 40 | 0.1583 | 2.9833 | 0.0257 |
| 21 | 0.2187 | 2.9229 | 0.0501 | 41 | 0.1563 | 2.9853 | 0.0250 |
| 22 | 0.2136 | 2.9280 | 0.0477 | 42 | 0.1545 | 2.9871 | 0.0244 |
| 23 | 0.2089 | 2.9327 | 0.0456 | 43 | 0.1526 | 2.9889 | 0.0238 |
| 24 | 0.2045 | 2.9371 | 0.0436 | 44 | 0.1509 | 2.9907 | 0.0233 |
| 25 | 0.2003 | 2.9413 | 0.0417 | 45 | 0.1492 | 2.9924 | 0.0227 |
| 26 | 0.1964 | 2.9452 | 0.0401 | 46 | 0.1476 | 2.9940 | 0.0222 |
| 27 | 0.1927 | 2.9488 | 0.0385 | 47 | 0.1460 | 2.9956 | 0.0217 |
| 28 | 0.1893 | 2.9523 | 0.0371 | 48 | 0.1445 | 2.9971 | 0.0213 |
| 29 | 0.1860 | 2.9556 | 0.0358 | 49 | 0.1430 | 2.9986 | 0.0208 |
|  |  |  |  | 50 | 0.1415 | 3.0001 | 0.0204 |

Explanation

Let     $P$ denote the proportion of items possessing a definite characteristic in an infinite population

$p$ denote the proportion of items possessing this characteristic in a sample of $n$ items drawn independently at random from the population

Then $\varphi_B(p) = 2 \arcsin \sqrt{p} = \arccos (1 - 2p)$, $0 < p < 1$, $0 < \varphi < \pi$

$\varphi_B(0) = 2 \arcsin \sqrt{1/4n}$; $\varphi_B(1) = \pi - \varphi_B(0)$

and   $V_{0.5}(\varphi_B)$ denotes the variance of $\varphi_B$ when $P = 0.5$.

*Note:* When $n > 50$, $\varphi_B(0) = \sqrt{1/4n}$, $\varphi_B(1) = \pi - \varphi_B(1)$, and $V_{0.5}(\varphi_B) = 1/n + 1/n^2$ are correct to within 1 unit in the fourth decimal place.

near 0.05 or 0.95, than the simpler adjustments $y(0,n)$ and $y(n,n)$. It is recommended, therefore, that, when special accuracy is desired, the values of $\varphi_B(0)$ and $\varphi_B(1)$ given in Table 16.4 of the present chapter be used in place of the values of $y(0,n)$ and $y(n,n)$ shown in the table of the $y$ transformation.

From Figs. 16.1 to 16.3 it is clear that, for a fairly wide range of values of $P$ about $P = 0.5$, the variance of $\varphi_B$, $nV(\varphi_B)$ being plotted as $V_3$ in these figures, is quite close to its value for $P = 0.5$, say $V_{0.5}(\varphi_B)$, and that $V_{0.5}(\varphi_B)$ approaches $1/n$ as $n$ increases. Indeed, from a consideration of these three cases ($n = 10, 20, 30$) it appears that

$$V_{0.5}(\varphi_B) = \frac{1}{n} + \frac{0.974}{n^2} + \frac{3.08}{n^3} - \frac{2.4}{n^4} \tag{11}$$

Using this formula, values of $V_{0.5}(\varphi_B)$ were computed for intermediate values of $n$, that is, for $10 < n < 20$ and $20 < n < 30$, and, in the absence of any other formula giving correct values for $n = 10, 20,$ and 30, this formula was used to compute values of $V_{0.5}(\varphi_B)$ for $n$ from 31 to 50. These values are tabulated, together with the known values for $n = 10, 20,$ and 30, in Table 16.4 and provide, it is believed, a reasonably accurate picture of the convergence of $V(\varphi_B)$ toward its theoretical value of $1/n$. The values of $V_{0.5}(\varphi_B)$ given in Table 16.4 are slightly larger than the values of the variance of $y$ given in the table of the $y$ transformation referred to above, but for $n \geq 25$ the excess is very slight, being roughly 1 unit or less in the fourth decimal place.

Fisher and Yates have suggested that, when special accuracy is desired, observed angles be replaced by adjusted angular values in much the same manner as observed probits are replaced by adjusted probits. They have given a table to facilitate such adjustments.* The adjustments proposed by Fisher and Yates are functions of the actual proportions observed in a given instance and thus are of a wholly different type from Bartlett's adjustment, which may be regarded as a convenient rule of thumb. The theoretical basis of the adjustments advocated by Fisher and Yates has been given in a recent paper by Cochran.†

Cochran has given an excellent discussion‡ of the presentation of results of statistical analyses in which transformations of original data have been employed to facilitate their statistical analysis. Curtiss§ has provided a general mathematical theory of transformations of data to render their analysis more tractable.

---

* FISHER and YATES, *Statistical Tables*, Table XIV.

† COCHRAN, "The Analysis of Variance When Experimental Errors Follow the Poisson or Binomial Laws," p. 342.

‡ COCHRAN, "Analysis of Replicated Experiments," p. 168 and p. 170.

§ CURTISS, "On Transformations."

In conclusion, it must be remembered that while in large samples the variance of a transformed proportion is independent of the true proportion for most practical purposes, the variance of the transformed value $\varphi$ is still proportional to $1/n$, where $n$ is the number of observations upon which the observed proportion is based. Consequently, if $p_1, p_2, p_3, \ldots$ are proportions based on different numbers of observations $n_1, n_2, n_3, \ldots$ and if the $n$'s differ widely, then the variances of the corresponding angular values may be so unequal, because of the variation of the $n$'s, that the advantages of the transformation are lost almost entirely. Similarly, the advantages of the inverse sine transformation may be lost to a large extent when the data to be analyzed are subject to other forms of variability in addition to sampling variation so that the variance of an observed proportion $p$ is given by $P(1 - P)/n + \sigma_e^2$, where the first term represents the variation of simple sampling and the second term the extraneous variation. In a recent paper* Cochran has discussed the analysis of fractions or percentages based on different numbers of observations $(a)$ when extraneous variation is absent and $(b)$ when extraneous variation is present.

## 4. TECHNICAL NOTES

### 4.1. General Transformation to Stabilize Variance

**4.1.1. Basic Theorem.**—The transformation of random variables to stabilize variance may be based upon the following theorem:

Let $\theta$, $(\theta_1 \leq \theta \leq \theta_2)$, be a parameter of a statistical universe (population), and let $T$ be an estimator of $\theta$, $[T_1(\theta) \leq T \leq T_2(\theta)]$, based on $n$ observations from the universe such that

$$\left. \begin{array}{r} \text{Expected value of } T \equiv E(T) = \theta \\[1em] \text{Variance of } T \equiv V(T) \equiv E(T - \theta)^2 = \dfrac{g(\theta)}{n} \\[1em] \text{Third moment of } T \text{ about } \theta \equiv \mu_3(T) \equiv E(T - \theta)^3 \\[1em] = o\left(\dfrac{1}{n}\right) \text{ as } n \to \infty \, \dagger \\[1em] \text{Fourth moment of } T \text{ about } \theta = \mu_4(T) \equiv E(T - \theta)^4 \\[1em] = o\left(\dfrac{1}{n}\right) \text{ as } n \to \infty \end{array} \right\} \quad (12)$$

* Cochran, "Analysis of Variance for Percentages Based on Unequal Numbers."

† By $\varphi(n) = o(n^k)$ as $n \to \infty$ we mean that $\lim\limits_{n \to \infty} \varphi(n)/n^k = 0$. In particular, $o(1)$ means a function that tends to zero as $n \to \infty$.

By $\varphi(n) = O(n^k)$ as $n \to \infty$ we mean that $|\varphi(n)| < K \cdot n^k$ for $n \geq n_o$, where $K$ is finite and does not depend upon $n$, and $n_o$ is some (large) value of $n$. In particular, $O(1)$ means a function that is bounded as $n \to \infty$.

This notation is in agreement with that introduced by E. Landau in 1909 and is now well established.

If the function

$$f(T) = \int \frac{dT}{\sqrt{g(T)}} \tag{13}$$

and its first two derivatives exist for all admissible values of $T$, except perhaps for a set of measure zero, and if the mean value of its second derivative and of the square of the second derivative are finite for $\theta_1 < \theta < \theta_2$ and every value of $n$, that is, if

$$E[f^{(2)}(T)] = O(1)$$

and

$$E[f^{(2)}(T)]^2 = O(1) \tag{14}$$

as $n \to \infty$ for $\theta_1 < \theta < \theta_2$ and $T_1(\theta) \leq T \leq T_2(\theta)$, then

$$V[f(T)] = \frac{1}{n} + o\left(\frac{1}{n}\right) \quad \text{as} \quad n \to \infty \tag{15}$$

The first of assumptions (12) implies that $T$ is an unbiased estimator of $\theta$. The three further assumptions are introduced for the purposes of the theorem and are not very restrictive. Indeed, assumptions (12) are generally satisfied when $T$ is some sort of average that has been adjusted so as to be an unbiased estimator of a population parameter of interest.

**4.1.2. Proof of Basic Theorem.**—Let $f(T)$ be a function of $T$ such that its second derivative $f^{(2)}(T)$ exists for $T_1(\theta) < T < T_2(\theta)$, except perhaps for a set of points of measure zero, where $T_1(\theta) \leq T \leq T_2(\theta)$ is the admissible range of values of $T$, that is, $P[T_1(\theta) \leq T \leq T_2(\theta)|\theta] = 1$. Since $E(T) \equiv \theta$ by assumption, it follows that $\theta$ lies in the admissible range of values for $T$. Hence $f(T)$ may be expressed as follows, with the aid of Taylor's formula and Lagrange's form of the remainder term:

$$f(T) = f(\theta) + \frac{(T - \theta)}{1!} f^{(1)}(\theta) + \frac{(T - \theta)^2}{2!} f^{(2)}(t) \tag{16}$$

where $t = \theta + \xi_T(T - \theta)$, $0 < \xi_T < 1$, and the notation $\xi_T$ is used to emphasize the fact that the value of $\xi$ depends upon the value of $T$ concerned, $T_1(\theta) \leq T \leq T_2(\theta)$.

Taking the mathematical expectation of each side of equation (16), it is found that

$$E[f(T)] = f(\theta) + \frac{V(T)}{2!} \cdot E[f^{(2)}(t)] \tag{17}$$

$$= f(\theta) + o(1) \quad \text{as } n \to \infty$$

by virtue of assumption (14) and the second of assumptions (12). This shows that these assumptions, together with those regarding the existence and differentiability of $f(T)$, are sufficient to ensure that, as $n$ increases without limit, the expected value of the function $f(T)$ converges

stochastically to $f(\theta)$, the value of the function when $T$ assumes its expected value.

Subtracting the left and right sides of (17) from the left and right sides, respectively, of (16), squaring, and taking expected values, it is found that

$$V[f(T)] \equiv E\{f(T) - E[f(T)]\}^2 \qquad\qquad (18)$$
$$= [f^{(1)}(\theta)]^2 \cdot V(T) + \frac{3E^2[f^{(2)}(t)]V^2(T)}{4} + \frac{E[f^{(2)}(t)]^2\mu_4(T)}{4}$$
$$+ \frac{f^{(1)}(\theta) \cdot E[f^{(2)}(t)]\mu_3(T)}{1}$$

In view of the assumptions (12) and (14), this becomes

$$V[f(T)] = [f^{(1)}(\theta)]^2 \cdot V(T) + o\left(\frac{1}{n}\right) \qquad \text{as } n \to \infty \qquad (19)$$

$$= [f^{(1)}(\theta)]^2 \cdot \frac{g(\theta)}{n} + o\left(\frac{1}{n}\right) \qquad \text{as } n \to \infty$$

Consequently, for the function defined by equation (13), equation (19) reduces to equation (15). This completes the proof.

Curtiss* has considered somewhat more fully the mathematical aspects of transformations of random variables to simplify their variances, and his paper should be consulted by persons interested in a comprehensive mathematical treatment.

### 4.2. Applications of the General Procedure

**4.2.1. Proportion of Items Possessing a Particular Attribute in a Sample.**—Let $\theta = P$, the proportion of items in an infinite population that possess a particular characteristic $A$, and let $T = p$, the proportion of items possessing $A$ in a sample of $n$ items drawn independently at random from the population. It is well known that

$$E(p) = P$$
$$V(p) = \frac{P(1 - P)}{n}$$
$$\mu_3(p) = \frac{P(1 - P)(1 - 2P)}{n^2} \qquad\qquad (20)$$
$$\mu_4(p) = \frac{3P^2(1 - P)^2}{n^2} + \frac{P(1 - P)(1 - 6P + 6P^2)}{n^3}$$

Clearly $\mu_3(p)$ and $\mu_4(p)$ are both $o(1/n)$ as $n \to \infty$, and $g(P) = P(1 - P)$. Consequently, if there is a transformation that will stabilize the variance, the theorem implies that it will be

$$f(p) = \int \frac{dp}{\sqrt{p(1 - p)}} = 2 \arcsin \sqrt{p} + C = \arccos (1 - 2p) + C \qquad (21)$$

---

\* CURTISS, "On Transformations Used in the Analysis of Variance."

where $C$ is an arbitrary constant, which may conveniently be taken as 0. This yields the transformation $\varphi(p)$, defined in equation (4), $\varphi(p)$ being expressed in radian measure. It is not difficult to show that the second derivative of $\varphi(p)$ satisfies the requirements of the transformation theorem for $0 < P < 1$, whence

$$V[\varphi(p)] = \frac{1}{n} + o\left(\frac{1}{n}\right) \qquad \text{as } n \to \infty \tag{22}$$

**4.2.2. Mean of a Sample from a Poisson Exponential Population.—** Let $\theta = \mu$, the parameter (mean) of a Poisson exponential population, and let $T = \bar{x}$, the mean of a random sample of $n$ independent observations from this population. It is well known that

$$E(\bar{x}) = \mu$$
$$V(\bar{x}) = \frac{\mu}{n}$$
$$\mu_3(\bar{x}) = \frac{\mu}{n^2} \tag{23}$$
$$\mu_4(\bar{x}) = \frac{3\mu^2}{n^2} + \frac{\mu}{n^3}$$

Clearly $\mu_3(\bar{x})$ and $\mu_4(\bar{x})$ are both $o(1/n)$ as $n \to \infty$, and $g(\mu) = \mu$. Consequently, the theorem implies that, if there exists a transformation that will stabilize the variance, it will be

$$f(\bar{x}) = \int \frac{d\bar{x}}{\sqrt{\bar{x}}} = 2\sqrt{\bar{x}} + C \tag{24}$$

It is convenient to take $C = 0$. Bartlett* chose to drop the 2 also and recommended the transformation $y = \sqrt{\bar{x}}$. The first and second derivatives of $y$ with respect to $\bar{x}$ are

$$y' = \frac{1}{2\sqrt{\bar{x}}} \qquad y'' = \frac{-1}{4(\bar{x})^{3/2}} \tag{25}$$

and

$$E(y') = \frac{1}{2\sqrt{\mu}} \text{ approximately}$$
$$E(y'') = \frac{-1}{4(\mu)^{3/2}} \text{ approximately} \tag{26}$$
$$[E(y'')]^2 = \frac{1}{16\mu^3} \text{ approximately}$$

---

* BARTLETT, "Square Root Transformation."

Hence, equation (18) above becomes for $y = \sqrt{\bar{x}}$

$$V(y) = \frac{1}{4n} + \frac{3}{64\mu n^2} + \frac{1}{64}\left(\frac{3}{\mu n^2} + \frac{1}{\mu^2 n^3}\right) - \frac{1}{8\mu n^2} \text{ approximately} \quad (27)$$
$$= \frac{1}{4n} - \frac{1}{32\mu n^2} + \frac{1}{64\mu^2 n^3}$$

Therefore, as $n \to \infty$ for fixed $\mu$

$$V(y) = \frac{1}{4n} + o\left(\frac{1}{n}\right) \quad (28)$$

and as $\mu \to \infty$ for fixed $n$

$$V(y) = \frac{1}{4n} + o\left(\frac{1}{\mu}\right) \quad (29)$$

Consequently, if either $n$ or $\mu$ is large, $V(y) = 1/4n$, approximately.

**4.2.3. Estimate of the Variance of a Normal Population Based on $n$ Degrees of Freedom.**—Let $\theta = \sigma^2$, the variance of a normal population, and let $T = s^2$, an estimate of $\sigma^2$ such that $ns^2/\sigma^2$ is distributed as $\chi^2$ for $n$ degrees of freedom. It is not difficult to show that

$$
\begin{aligned}
E(s^2) &= \sigma^2 \\
V(s^2) &= \frac{2\sigma^4}{n} \\
\mu_3(s^2) &= \frac{8\sigma^6}{n^2} \\
\mu_4(s^2) &= \frac{12(n+4)\sigma^8}{n^3}
\end{aligned}
\quad (30)
$$

Clearly $\mu_3$ and $\mu_4$ are $o(1/n)$ as $n \to \infty$ and $g(\sigma^2) = (\sigma^2)^2$. Therefore, the theorem implies that, if a transformation that will stabilize the variance exists, it will be

$$f(s^2) = \int \frac{ds^2}{\sqrt{2}\,s^2} = \frac{\log_e s^2}{\sqrt{2}} + C = \sqrt{2}\log_e s + C \quad (31)$$

For convenience, we may take $C = 0$ and consider the transformation $y = \log_e s^2 = 2.30259 \log_{10} s^2$. Since

$$
\begin{aligned}
E(y') &= \frac{1}{\sigma^2} \text{ approximately} \\
E(y'') &= -\frac{1}{\sigma^4} \text{ approximately} \\
[E(y'')]^2 &= \frac{1}{\sigma^8} \text{ approximately}
\end{aligned}
\quad (32)
$$

the type of reasoning employed in connection with the preceding example shows that

$$V(\log_e s^2) = \frac{2}{n} + o\left(\frac{1}{n}\right) \qquad \text{as } n \to \infty \text{ for fixed } \sigma^2 \qquad (33)$$

$$V(\log_e s^2) = \frac{2}{n} + o\left(\frac{1}{\sigma^2}\right) \qquad \text{as } \sigma^2 \to \infty \text{ for fixed } n \qquad (34)$$

Consequently, $V(\log_e s^2) = 2/n$, approximately, provided either $n$ or $\sigma^2$ is sufficiently large.

**4.2.4. Correlation Coefficient Computed from a Sample from a Normal Population.**—Let $\theta = \rho$, the product moment coefficient of correlation in a bivariate normal population, and let $T = r$, a correlation coefficient computed from a sample of $n$ observations drawn independently at random from the population. The mean, variance, and higher moments of $r$ are exceedingly complicated functions of $\rho$ and $n$, but when $n$ is large and $|\rho| \ll 1$,

$$
\begin{aligned}
E(r) &= \rho \text{ approximately} \\
V(r) &= \frac{(1 - \rho^2)^2}{n} \text{ approximately} \\
\mu_3(r) &= o\left(\frac{1}{n}\right) \qquad \text{as } n \to \infty \\
\mu_4(r) &= o\left(\frac{1}{n}\right) \qquad \text{as } n \to \infty
\end{aligned}
\qquad (35)
$$

Consequently, the theorem suggests the use of the following transformation to stabilize the variance:

$$f(r) = \int \frac{dr}{1 - r^2} = \frac{1}{2} \int \left(\frac{1}{1 - r} + \frac{1}{1 + r}\right) dr = \frac{1}{2} \log_e \frac{1 + r}{1 - r} + C \quad (36)$$

Taking $C = 0$, this yields

$$z = \frac{1}{2} \log_e \frac{1 + r}{1 - r} = \tanh^{-1} r \qquad (37)$$

the transformation of $r$ introduced by R. A. Fisher.*

**4.3. A Method of Discovering Bartlett's Adjustment**

Let the values of

$$\varphi(p) = 2 \arcsin \sqrt{p} \qquad (38)$$

---

* FISHER, "On the 'Probable Error' of a Coefficient of Correlation Deduced from a Small Sample," *Metron*, Vol. 1 (1921), pp. 1–32.

where

$$p = \frac{x}{n} \qquad (x = 0, 1, 2, \cdots, n) \qquad (39)$$

be plotted on the arithmetic scale of arithmetic-probability paper against the cumulative probability

$$\sum_{r=0}^{x} \frac{n!}{r!(n-r)!} P^r (1-P)^{n-r} \qquad (40)$$

plotted on the probability scale, for a particular value of $n$ and a selected value of $P$. If $n \geq 10$ and $0.05 \leq P \leq 0.95$, it will be found that the plotted points for $2 \leq x \leq n - 2$ lie very nearly on a straight line, that the points for $x = 1$ and $x = n - 1$ depart slightly from the extension of this line, but that the points for $x = 0$ and $x = n$ lie far off the extension of this line. If the values $\varphi(0) = 0$ and $\varphi(1) = \pi$ are replaced by scores $\varphi(0) = a_n$ and $\varphi(1) = b_n$ such that the corresponding plotted points fall on a gently curving extrapolation of the curve through the plotted points for $2 \leq x \leq n - 2$, it will be found that

$$a_n = 2 \arcsin \sqrt{\frac{1}{4n}} \text{ approximately}$$
$$b_n = \pi - a_n \text{ approximately} \qquad (41)$$

The foregoing suggests that a better adjustment than Bartlett's could be obtained by extending in both directions the line of best fit through the points for $2 \leq x \leq n - 2$ and using this line to define the scores for $x = 0, 1, n - 1,$ and $n$. This was tried and some improvement was noted, but for $n \geq 10$ and $0.05 \leq P \leq 0.95$ the improvement did not appear to be worth the trouble.

This procedure may be found useful for obtaining adjustments of the Bartlett type in connection with other transformations to stabilize variance.

### 5. REFERENCES*

**5.1. Inverse Sine Transformation**

BARTLETT, M. S.: "Square Root Transformation in Analysis of Variance," *Supplement to the Journal of the Royal Statistical Society*, Vol. 3 (1936), pp. 68–78.

* As this volume went to press, binomial-probability graph paper utilizing the inverse sine transformation was put out by the Codex Book Company, Inc., Norwood, Mass. The paper was designed by Frederick Mosteller and John W. Tukey, pursuant to a suggestion published in 1943 by R. A. Fisher and K. Mather (*Annals of Eugenics*, Vol. 12, p. 13). Messrs. Mosteller and Tukey plan to write an explanation of its design and uses, probably for the *Journal of the American Statistical Association*. In the meantime, they have prepared a brief statement that may be obtained from the Codex Book Company.

The paper may be ordered in two weights, heavy (no. 32,298) and light (no. 31,299). The sheets measure 8½ by 11 in.

———: "Subsampling for Attributes," *Supplement to the Journal of the Royal Statistical Society*, Vol. 4 (1937), pp. 131–135.   Table II, p. 132, gives $y = \sin^{-1} \sqrt{x}$ in radians to 4 decimals for $x = 0.01$ (0.01) 0.50.

———: "Some Examples of Statistical Methods of Research in Agriculture and Applied Biology," *Supplement to the Journal of the Royal Statistical Society*, Vol. 4 (1937), pp. 137–183.

BATEN, WILLIAM DOWELL, and HENDERSON, EARL WILTON: "A Method of Handling Ratios by the Analysis of Variance," *Poultry Science*, Vol. 20 (1941), pp. 227–231.

BLISS, C. I.: "The Analysis of Field Experimental Data Expressed in Percentages," *Plant Protection*, No. 12 (1937), pp. 67–77.   [In Russian, with English subtitles and summary.]   Table 1, pp. 70–72, gives arcsin $\sqrt{P/100}$ in degrees to 2 decimals for (percentages) $P = 0.00$ (0.01) 1.00 (0.1) 99.00 (0.01) 100.00.

———: "The Transformation of Percentages for Use in the Analysis of Variance," *Ohio Journal of Science*, Vol. 38 (1938), pp. 9–12.   Table I, p. 12, gives arcsin $\sqrt{P/100}$ in degrees to 1 decimal place for (percentages) $P = 0.0$ (0.1) 1.0.

CLARK, ANDREW, and LEONARD, WARREN H.: "The Analysis of Variance with Special Reference to Data Expressed in Percentages," *Journal of the American Society of Agronomy*, Vol. 31 (1939), pp. 55–66.

COCHRAN, W. G.: "Some Difficulties in the Statistical Analysis of Replicated Experiments," *Empire Journal of Experimental Agriculture*, Vol. 6 (1938), pp. 157–175.

———: "The Analysis of Variance When Experimental Errors Follow the Poisson or Binomial Laws," *Annals of Mathematical Statistics*, Vol. 11 (1940), pp. 335–347.

———: "Analysis of Variance for Percentages Based on Unequal Numbers," *Journal of the American Statistical Association*, Vol. 38 (1943), pp. 287–301.

CURTISS, J. H.: "On Transformations Used in the Analysis of Variance," *Annals of Mathematical Statistics*, Vol. 14 (1943), pp. 107–122.

FISHER, R. A.: "On the Dominance Ratio," *Proceedings of the Royal Society of Edinburgh*, Vol. 42 (1921–1922), pp. 321–341.

———: "The Distribution of Gene Ratios for Rare Mutations," *Proceedings of the Royal Society of Edinburgh*, Vol. 50 (1929–1930), pp. 205–220.

———, and YATES, FRANK: *Statistical Tables for Biological, Agricultural and Medical Research*, 2nd ed., Oliver and Boyd, Ltd., London and Edinburgh, 1943, 98 + viii pp.   Table XII gives arcsin $\sqrt{P/100}$ in degrees to 1 decimal place for (percentages) $P = 0$ (1) 99.   Table XIII gives arcsin $\sqrt{p}$ in degrees to 1 decimal place for (fractions) $p = x/n$ corresponding to $n = 2$ (1) 30 and integral values of $x$ from 1 to the largest integer less than $n/2$.   Table XIV gives "angular values for adjustments of special accuracy," namely, $\varphi + \frac{1}{2} \cot \varphi$ and cosec $2\varphi$, to 1 decimal place for $\varphi = 45$ (1) 89 degrees.

FULCHER, JOHN S., and ZUBIN, JOSEPH: "The Item Analyzer: A Mechanical Device for Treating the Four Fold Table in Large Samples," *Journal of Applied Psychology*, Vol. 26 (1942), pp. 511–522.

SNEDECOR, GEORGE W.: *Statistical Methods Applied to Experiments in Agriculture and Biology*, 4th ed., The Iowa State College Press, Ames, 1946.   Table 16.8 is a reproduction of the table published by C. I. Bliss in *Plant Protection*.

Statistical Research Group, Columbia University, *Sequential Analysis of Statistical Data; Applications*, rev. ed., Columbia University Press, New York, 1945.   Table 3.64, opposite p. 3.56 of Sec. 3, gives $y(s,n) = 2 \arcsin \sqrt{s/n}$ in radians to 4 decimals for $n = 10$ (1) 50 and $s = 1$ (1) $n - 1$, together with values of $y(0,n) = \sqrt{2/n}$ and $y(n,n) = 3.1416 - \sqrt{2/n}$ to 4 decimals and values of $V = 1/n + 1/n^2$ to 6 decimals, for $n = 10$ (1) 50.

ZUBIN, JOSEPH: "Note on a Transformation Function for Proportions and Percentages," *Journal of Applied Psychology*, Vol. 19 (1936), pp. 213–220. Table 4, p. 220, gives $T = 2$ arcsin $\sqrt{p}$ in radians to 4 decimals for (proportions) $p = 0.00$ (0.01) 0.99.

———, "Note on a Graphic Method for Determining the Significance of the Difference between Group Frequencies," *Journal of Educational Psychology*, Vol. 27 (1936), pp. 431–444.

———: "Nomographs for Determining the Significance of the Differences between the Frequencies of Events in Two Contrasted Series or Groups," *Journal of the American Statistical Association*, Vol. 34 (1939), pp. 539–544.

**5.2. Mathematical Tables**

HAYASHI, KEIICHI: *Sieben-und mehrstellige Tafeln der Kreis-und Hyperbelfunktionen*, Julius Springer, Berlin, 1926, 283 + vi pp.

PEARSON, KARL: *Tables of the Incomplete Beta Function*, University Press, Cambridge, England, 1934, 494 + lix pp. ·

SALZER, H. E., and ISAACSON, E.: "Mathematical Tables—Errata, No. 35," *Mathematical Tables and Other Aids to Computation*, Vol. 1 (1943–1945), pp. 196–197. A list of errors in Hayashi's table of arcsin $x$.

# CHAPTER 17

# ELEMENTS OF SEQUENTIAL ANALYSIS

This chapter was prepared by
W. Allen Wallis, University of Chicago

# CHAPTER 17

## ELEMENTS OF SEQUENTIAL ANALYSIS*

### 1. INTRODUCTION

### 1.1. Nature of Sequential Tests

A sequential test of a statistical hypothesis consists of making a certain calculation after each observation (or group of observations) and deciding on the basis of this calculation whether (1) to accept the hypothesis under test, (2) to accept the alternative to that hypothesis, or (3) to suspend judgment pending the examination of more data. The criteria for these decisions are such that the hypothesis is accepted or rejected as soon as, but not before, the data are sufficient for doing so with predetermined degrees of reliability.

### 1.2. Advantages and Disadvantages

Sequential is superior to nonsequential analysis whenever (1) the data become available serially and (2) the cost of the data (in terms of time, labor, material, and so forth) is more or less proportional to the amount of data. Nonsequential analysis is superior whenever (1) the amount of data is fixed or (2) the cost of the data is largely overhead, hence more or less independent of the amount of data. *Superiority* consists in minimizing a set of quantities $(N,\alpha,\beta)$, where $N$ is the number of observations, $\alpha$ is the risk (that is, the probability) of erroneously rejecting the hypothesis under test, and $\beta$ is the risk of erroneously accepting this hypothesis. Sequential and nonsequential tests differ in the constraints under which this set is minimized. Nonsequential tests treat $N$ as fixed and are so designed that with either risk $\alpha$ or $\beta$ fixed the other is minimized. Sequential tests treat the number of observations $N$

---

* Sequential analysis was developed in the Statistical Research Group in 1943, the theory principally by A. Wald and the techniques of application by numerous members of the group, including Kenneth J. Arnold, H. A. Freeman, Milton Friedman, M. A. Girshick, Edward Paulson, A. Wald, and W. Allen Wallis.

This chapter is an extension of a memorandum by the author of the chapter, "Sequential Tests of Statistical Significance," which is cited as "SRG 180" in Appendix B of *Sequential Analysis*. A more extensive exposition of the theory than is given here will be found in Appendix B of *Sequential Analysis*, and a more extensive summary of methods of application than is given here will be found in "Sequential Tests." For a complete and rigorous account of the theory, see "Sequential Tests"; for a full account of methods of application and interpretation, including computational aids, see *Sequential Analysis*. (For full citations, see Sec. 8.)

as a variable and are so designed that, for fixed risks $\alpha$ and $\beta$, the expected (that is, average) number of trials $\bar{N}$ required to reach a decision is minimized.    If for a nonsequential test $N$ is made large enough so that, with $\alpha$ fixed, $\beta$ will not exceed a predetermined amount, this value of $N$ will exceed (frequently by as much as 100 percent) the $\bar{N}$ required with a sequential test for the same $\alpha$ and $\beta$.    Thus, when $N$ is readily subject to variation, sequential tests are superior; when $N$ is not readily varied, nonsequential tests are superior.*

### 1.3.  General Procedure

The basic quantity computed after each observation in a sequential test is known in statistical literature as the "likelihood ratio." The probability of obtaining the observations actually at hand is computed first on the assumption that the hypothesis under test (designated the "null hypothesis") is true, and second on the assumption that the alternative to this hypothesis (the "alternative hypothesis") is true.    If these probabilities are designated by $P_0$ and $P_1$, respectively, the likelihood ratio is simply $P_1/P_0$.    In a sequential test it is computed after each observation (or group of observations), each computation being based on all the observations available.    If the likelihood ratio ever reaches a certain level $A$ ($A > 1$), the collection of data (experiment) halts and the alternative hypothesis is accepted.    If the likelihood ratio ever falls to a certain level $1/B$ ($B > 1$), the experiment ceases and the null hypothesis is accepted.    When the likelihood ratio is between $A$ and $1/B$, the evidence is insufficient for either decision with the required reliability.    The required reliability, that is, the $\alpha$ and $\beta$ of the preceding paragraph, completely determines $A$ and $B$; in fact, $A = (1 - \beta)/\alpha$ and $B = (1 - \alpha)/\beta$.

Since in most cases the successive observations are independent, the likelihood ratio for a set of observations is simply the likelihood ratio for the last observation multiplied by that for the set of preceding observations.    Consequently, using the logarithm of the likelihood ratio reduces the arithmetic to addition and subtraction.    After each observation, simply add $\log P_1$ to and subtract $\log P_0$ from the algebraic sum of such quantities for all preceding observations.    If this sum ever reaches

$$a = \log A = \log \frac{1 - \beta}{\alpha} \tag{1}$$

accept the alternative hypothesis; if it ever falls to $-b$, where

* A detailed discussion of the conditions under which sequential analysis is advantageous and those under which it is disadvantageous is given in Statistical Research Group, *Sampling Inspection*, pp. 33–38 and 91–97.    Although the discussion there is in terms of specific circumstances, its generalization is obvious.

$$b = \log B = \log \frac{1 - \alpha}{\beta} \qquad (2)$$

accept the null hypothesis.   Otherwise, continue the experiment.

### 1.4. Simplifications

The preceding paragraph is applicable whatever the null and alternative hypotheses may be, provided only that all parameters of each hypothesis are uniquely specified.   In particular problems, the procedure can usually be simplified by expressing $\log P_1$ and $\log P_0$ directly in terms of the observations, then adjusting the terminat. on levels to simplify the expression.   This is illustrated below for three common tests: (1) for the probability of success in a binomial distribution; (2) for the mean in a normal distribution; and (3) for the difference in probability of success between two binomial distributions.   The arithmetic can be reduced in the first and third cases to mere counting and in the second case to simple addition.

### 2. BINOMIAL DISTRIBUTION

### 2.1. Algebraic Simplification of General Procedure

Suppose it is required to determine whether the proportion $p$ of successes, from a certain type of operation that results in either success or failure, is $p_0$.   The alternative against which this is to be tested is that $p = p_1$ $(p_1 > p_0)$.   The value of $p_0$ may be the minimum acceptable proportion for the operation, it may be a value observed from long experience with similar operations, or it may have been deduced from theoretical reasoning.   The value of $p_1$ may be the least excess over $p_0$ worth detecting, the proportion expected on theoretical grounds, and so forth.   Suppose that, after $N$ operations, $S$ successes have occurred. The probability of this happening (in the sequence actually observed) is $p_1^S(1 - p_1)^{N-S}$ under the alternative hypothesis and $p_0^S(1 - p_0)^{N-S}$ under the null hypothesis.   The likelihood ratio is therefore

$$\left(\frac{p_1}{p_0}\right)^S \left[\frac{(1 - p_1)}{(1 - p_0)}\right]^{N-S}$$

and its logarithm is

$$S \log \left(\frac{p_1}{p_0}\right) + (N - S) \log \left[\frac{(1 - p_1)}{(1 - p_0)}\right]$$

The hypothesis that $p = p_0$ is to be rejected, and the hypothesis that

$p = p_1$ accepted, if this is ever as great as $a$, that is, if $S$ ever reaches

$$S_1 = \frac{a + N \log\left(\dfrac{1 - p_0}{1 - p_1}\right)}{\log\left[\dfrac{p_1(1 - p_0)}{p_0(1 - p_1)}\right]} \tag{3}$$

This is simply a linear function of $N$ with positive intercept. Similarly, the hypothesis that $p = p_0$ is to be accepted if $S$ is ever as small as

$$S_0 = \frac{-b + N \log\left(\dfrac{1 - p_0}{1 - p_1}\right)}{\cdot \cdot \log\left[\dfrac{p_1(1 - p_0)}{p_0(1 - p_1)}\right]} \tag{4}$$

another (and parallel) linear function of $N$ with negative intercept.

### 2.2. Graphical Procedure

Thus, the simplest procedure is to prepare a chart with $N$ on the horizontal and $S$ on the vertical axis. On this chart the two parallel lines $S_1$ and $S_0$ are drawn. After each observation is obtained, the total number of successes observed is plotted against the total number of observations made. If, for any $N$, $S$ is as great as $S_1$, discontinue the experiment and accept the hypothesis that $p = p_1$. If, for any $N$, $S$ is as small as $S_0$, discontinue the experiment and accept the hypothesis that $p = p_0$. When $S$ is between $S_0$ and $S_1$, no decision is possible with a risk as small as $\alpha$ or $\beta$. A variation of this procedure is to plot the number of failures $F$ ($= N - S$) on the horizontal axis instead of $N$, retaining $S$ for the vertical scale. Then the two decision lines become, respectively,

$$S_1' = \frac{a + F \log\left(\dfrac{1 - p_0}{1 - p_1}\right)}{\log\dfrac{p_1}{p_0}} \tag{5}$$

and

$$S_0' = \frac{-b + F \log\left(\dfrac{1 - p_0}{1 - p_1}\right)}{\log\dfrac{p_1}{p_0}} \tag{6}$$

*Caution:* Be sure that the same logarithmic base is used for $a$ and $b$ as is used elsewhere in $S_1$ and $S_0$ or $S_1'$ and $S_0'$.

## 3. NORMAL DISTRIBUTION

### 3.1. Algebraic Simplification of General Procedure

Suppose the standard deviation of a normal (Gaussian) distribution is known to be $\sigma$ and that it is required to test the null hypothesis that the mean $m$ is $m_0$ against the alternative hypothesis that $m = m_1$ $(m_1 > m_0)$. If $N$ observations, $x_1, x_2, \ldots, x_N$, are obtained, the probability of this sequence is

$$(\sigma \sqrt{2\pi})^{-N} \exp \left\{ - \frac{1}{2\sigma^2} \left[ \sum_{i=1}^{N} (x_i - m_0)^2 \right] \right\} \qquad (7)$$

under the null hypothesis and the same expression, except with $m_1$ replacing $m_0$, under the alternative hypothesis. From these probabilities the *natural* logarithm of the likelihood ratio is found to be

$$\frac{m_1 - m_0}{\sigma^2} \left( \sum_{i=1}^{N} x_i - \frac{m_0 + m_1}{2} N \right) \qquad (8)$$

The hypothesis that $m = m_1$ is to be accepted if this is ever as great as $a$, that is, if $\Sigma x$ is ever as great as

$$M_1 = \frac{a\sigma^2}{m_1 - m_0} + \frac{m_1 + m_0}{2} N \qquad (9)$$

Similarly, the hypothesis $m = m_0$ is to be accepted if $\Sigma x$ is ever as small as

$$M_0 = \frac{-b\sigma^2}{m_1 - m_0} + \frac{m_1 + m_0}{2} N \qquad (10)$$

### 3.2. Graphical Procedure

As in the case of the binomial, prepare a chart with $N$ on the horizontal axis and $\Sigma x$ on the vertical axis, and draw on it the two parallel straight lines $M_1$ and $M_0$, of which $M_1$ has a positive and $M_0$ a negative intercept. As each observation is obtained, add it to the sum of all previous observations and plot the total against $N$. If the plotted point is ever as high as $M_1$, terminate the experiment and accept the hypothesis that $m = m_1$. If it is ever as low as $M_0$, terminate the experiment and accept the hypothesis that $m = m_0$. When it is between $M_0$ and $M_1$, no decision can be reached without a risk of error exceeding $\alpha$ or $\beta$.

*Caution:* Be sure that $a$ and $b$ are *natural* logarithms. (Natural logarithms are 2.3 times the corresponding common logarithms.)

## 4. DOUBLE DICHOTOMIES

### 4.1. Null and Alternative Hypotheses

Data from two different operations, each of which results in either success or failure, constitute a double dichotomy. To determine sequentially whether the proportions of successes are the same for the two operations (which is logically identical with testing whether the two operations are represented in the same proportion among the successes as among the failures), set up the null hypothesis that the ratio of successes to failures is the same for the two operations, and the alternative hypothesis that the ratio of successes to failures is $u$ ($u > 1$) times as great for the second operation as for the first. Thus, if $p_1$ and $p_2$ are the proportions of successes for the two operations, the null hypothesis is that $\frac{p_2/(1 - p_2)}{p_1/(1 - p_1)} = 1$, which is equivalent to $p_1 = p_2$. The alternative hypothesis is that $\frac{p_2/(1 - p_2)}{p_1/(1 - p_1)} = u$, which is equivalent to the relation $p_2 = up_1/[1 + p_1(u - 1)]$. Then collect observations in pairs, each pair containing one observation for each operation. Those pairs in which both operations succeed or both fail are disregarded, since they throw no light on the question of which operation yields the higher proportion of successes. Only the residual pairs, in which one and only one operation results in a success, are considered in the remainder of this section; these residual pairs will hereafter be called successes if the operation expected under the alternative hypothesis to be superior produces a success.

### 4.2. Algebraic Simplification of General Procedure

Under the null hypothesis, the proportion of such successes among the residual pairs will be $\frac{1}{2}$. Under the alternative hypothesis, the proportion of such successes will be $u/(1 + u)$. Thus, the problem reduces to that of the binomial distribution discussed in Sec. 2. If $S$ is the number of successes and if $F$ is the number of failures from $N$ residual pairs ($S + F = N$), the logarithm of the likelihood ratio may be written

$$S \log \frac{2u}{1 + u} + F \log \frac{2}{1 + u} \tag{11}$$

The alternative hypothesis (that the second operation is superior) is to be accepted if this is ever as great as $a$; this will occur if, for any $F$, $S$ is as much as

$$S_1 = \frac{a + F \log \frac{1 + u}{2}}{\log \frac{2u}{1 + u}} \tag{12}$$

a linear function of $F$ with positive intercept. Similarly, the null hypothesis (that the two operations are equally good) is to be accepted if the logarithm of the likelihood ratio is ever as small as $-b$, which will occur if, for any $F$, $S$ is as small as

$$S_0 = \frac{-b + F \log \frac{1 + u}{2}}{\log \frac{2u}{1 + u}} \tag{13}$$

a linear function of $F$, parallel to $S_1$ but with negative intercept.

## 4.3. Graphical Procedure

Thus, the simplest procedure is to prepare a chart on which the horizontal axis represents $F$, the number of residual pairs in which a failure has occurred with the operation expected under the alternative hypothesis to be superior, and the vertical axis represents $S$, the number of residual pairs in which that operation has succeeded. Draw on this chart the two parallel straight lines $S_0$ and $S_1$. After each residual pair is observed, plot a point whose coordinates represent the total number of successes and failures among all residual pairs obtained so far. If, for any value of $F$, $S$ is as great as $S_1$, halt the experiment and conclude that the second operation is superior. If, for any value of $F$, $S$ is as little as $S_0$, halt the experiment and conclude that the two operations are equally good. When $S$ is between $S_0$ and $S_1$, no conclusion can be reached without a risk of error in excess of $\alpha$ or $\beta$. An alternative procedure is to let the horizontal scale represent the number of residual pairs $N$ ($= S + F$), in which case the decision lines become, respectively,

$$S_1' = \frac{a + N \log \frac{1 + u}{2}}{\log u} \tag{14}$$

and

$$S_0' = \frac{-b + N \log \frac{1 + u}{2}}{\log u} \tag{15}$$

*Caution:* Be sure that the same logarithmic base is used for $a$ and $b$, as is used elsewhese in $S_0$ and $S_1$ or $S_0'$ and $S_1'$.

## 5. OTHER APPLICATIONS

### 5.1. Simple Problems

Obviously, the general procedure and the simplifications described in Secs. 1.3 and 1.4 can be applied to a wide range of problems. The application is straightforward whenever the problem is simple, in that a single parameter $\theta$ is under investigation and the purpose of the analysis is to discriminate between a null hypothesis that this parameter lies below [above] some level $\theta_0$ and an alternative hypothesis that it lies above [below] another level $\theta_1$, where $\theta_1 > \theta_0$ [$\theta_1 < \theta_0$]. For example, procedures for testing the mean of a Poisson distribution and for testing the standard deviation of a normal distribution are easily derived from the general procedure. More complicated problems can sometimes be reduced to a simple form for which the procedure is straightforward, as is illustrated by the case of double dichotomies.

### 5.2. Complex Problems

Sequential procedures have also been worked out successfully for certain more complicated problems. Important examples are (1) testing the null hypothesis that the mean of a normal distribution has a certain value $m_0$ against the two-sided alternative hypothesis that the mean is either greater than $m_1$ or less than $2m_0 - m_1$ ($m_1 > m_0$); (2) testing the same null and alternative hypotheses as in Sec. 3, but with no information about the standard deviation other than that given by the sequential sample (the sequential equivalent of Student's test); (3) testing the standard deviation of a normal distribution, with no information about the mean other than that given by the sequential sample; (4) testing the fraction of a normal distribution exceeding a certain level, with no information about the mean or standard deviation other than that given by the sequential sample.

## 6. OPERATING CHARACTERISTICS

### 6.1. General

Sequential tests are determined by four quantities, the values $\theta_0$ and $\theta_1$ specified by the null and alternative hypotheses for the parameter being investigated and the values $\alpha$ and $\beta$ of the two risks of error. For simple problems involving independent observations, for example, those described in Secs. 2, 3, and 4, the sequential-test procedure involves only three parameters, the slope $s$ and the two intercepts $-h_1$ and $h_2$ of the two decision lines. Obviously, then, many combinations of $\theta_0$, $\theta_1$, $\alpha$, and $\beta$ lead to the same sequential test. By finding these combinations, we can determine the probability that the test will lead to acceptance of the null hypothesis for any possible value of $\theta$, not merely $\theta_0$ and $\theta_1$.

The function (curve) giving the probability of acceptance $L_\theta$ for every possible value of $\theta$ is known as the operating-characteristic function (OC curve) of the test procedure; and $1 - L_\theta$ is known as the power function.

## 6.2. Five-point OC Curves

Although the equations of these curves are known, it ordinarily suffices to find five easily determined points on them. Two points are known from the fact that at the extreme values of $\theta$, for example, at $p = 0$ and $p = 1$ in the case of the binomial distribution, the probabilities of acceptance are 1 and 0. Two more points are known from the fact that the test was obtained in such a way that at $\theta_0$ and $\theta_1$ the probabilities of acceptance are $1 - \alpha$ and $\beta$. A fifth, and intermediate, point results from the fact that the probability of accepting the null hypothesis is

$$\frac{h_2}{h_1 + h_2}$$

(which is $\frac{1}{2}$ if $\alpha = \beta$) when $\theta$ is such that the expected value of each term in the cumulated sum is $s$, for example, when the parameter of a binomial distribution is $p = s$, when the mean of a normal distribution is $m = s$, or when the standard deviation of a normal distribution is $\sigma = \sqrt{s}$.

## 7. AVERAGE NUMBER OF OBSERVATIONS

## 7.1. General

The average number of observations $\bar{N}$ necessary to reach a decision will depend upon the risks of error and upon the divergence between the null and alternative hypotheses. $\bar{N}$ will be greater the smaller $\alpha$, the smaller $\beta$, or the smaller the divergence between the null and alternative hypotheses. Furthermore, it will have one value, $\bar{N}_0$, when the null hypothesis is true and another, $\bar{N}_1$, when the alternative hypothesis is true; in general its value $N_\theta$ will depend upon the true value $\theta$ of the parameter being tested. The function (curve) giving the average number of observations $\bar{N}_\theta$ for every possible value of $\theta$ is known as the average sample number function (ASN curve). Its equation may be written

$$\bar{N}_\theta = \frac{(1 - L_\theta)a - L_\theta b}{\bar{z}_\theta} \tag{16}$$

where $L_\theta$ is the probability, when the true value of the parameter is $\theta$, that the test will result in accepting the null hypothesis, and $\bar{z}_\theta$ represents the average value of the logarithm of the likelihood ratio when the true value of the parameter is $\theta$. In the case of the binomial distribution, for example, the likelihood ratio is either $p_1/p_0$ (when a success occurs),

or $(1 - p_1)/(1 - p_0)$ (when a failure occurs). If the true proportion of successes is $p$,

$$\dot{z}_p = p \log \frac{p_1}{p_0} + (1 - p) \log \left( \frac{1 - p_1}{1 - p_0} \right) \qquad (17)$$

$$= \log \frac{1 - p_1}{1 - p_0} + p \log \left[ \frac{p_1(1 - p_0)}{p_0(1 - p_1)} \right]$$

### 7.2. Five-point ASN Curves

Five points on the ASN curve are readily obtainable from the five known points on the OC curve. Two points are obtained from the extreme values of $\theta$ and the corresponding probabilities of acceptance, 1 and 0. Two more points, corresponding with $\theta_0$ and $\theta_1$, are obtained by taking $L_\theta$ equal to $1 - \alpha$ and $\beta$, respectively. The fifth, intermediate, point is obtained by multiplying together $h_1$ and $h_2$ and dividing the product by the variance of a term in the cumulated sum when $\theta$ is such that the expected value of each term is $s$; for example, the divisor of $h_1 h_2$ in testing the binomial proportion is $s(1 - s)$, in testing the mean of a Poisson distribution it is $s$, in testing the mean of a normal distribution of known standard deviation $\sigma$ it is $\sigma^2$, in testing the standard deviation of a normal distribution it is $2s^2$. This fifth value, $\bar{N}_s$, is usually (but not necessarily) near the maximum of the ASN curve.

For a given value of $\theta$, sampling fluctuations will cause $\bar{N}$ to vary from one sequential analysis to another, using the same values of $h_1$, $h_2$, and $s$. The distribution of $N$ is skewed in such a way that most samples require fewer than $\bar{N}_\theta$ observations; the variability is such that the sample size rarely exceeds $3\bar{N}_\theta$. For usual combinations of $\theta_0$, $\theta_1$, $\alpha$, and $\beta$, the ASN curve falls off sharply from its peak and $\bar{N}_0$ and $\bar{N}_1$ are substantially less than $\bar{N}_s$; this is not necessarily the case, however, and in fact it is possible for the maximum of the ASN curve to occur outside the interval $\theta_0$ to $\theta_1$, although $\theta = s$ is always between $\theta_0$ and $\theta_1$.

### 8. REFERENCES

FREEMAN, H. A., and OTHERS: *Sequential Analysis of Statistical Data: Applications.* SRG Report 255, submitted by the Statistical Research Group, Columbia University, to the Applied Mathematics Panel, National Defense Research Committee, July, 1944. Now unavailable.

Statistical Research Group, Columbia University: *Sampling Inspection*, McGraw-Hill Book Company, Inc., New York and London, 1948.

———: *Sequential Analysis of Statistical Data: Applications*, rev. ed., Columbia University Press, New York, 1945.

———: *Sequential Tests of Statistical Significance.* SRG Memorandum 180, submitted by the Statistical Research Group, Columbia University, to the Applied Mathematics Panel, National Defense Research Committee, April, 1944. Now unavailable.

WALD, A.: *Sequential Analysis*, John Wiley & Sons, Inc., New York, 1947, 212 + xii pp.

————: *Sequential Analysis of Statistical Data: Theory.*   SRG Report 75, submitted by the Statistical Research Group, Columbia University,  to the Applied Mathematics Panel, National Defense Research Committee, September, 1943.    Now unavailable.

————: "Sequential Tests of Statistical Hypotheses," *Annals of Mathematical Statistics*, Vol. 16 (1945), pp. 117–186.   Reprints obtainable from the Institute of Mathematical Statistics, 116 Rackham Hall, Ann Arbor, Michigan.

# INDEX*

* The references are to pages. The arrangement of major topics and symbols is alphabetical; subtopics are in alphabetical order, disregarding preliminary prepositions. Greek letters are alphabetized according to their word equivalents, e.g. α as "alpha," χ² as "chi-squared," and ξ as "xi." Certain symbols, for example, *A* and *a*, are used in a variety of meanings; in such cases each meaning is separately listed, and the arrangement is alphabetical with respect to first word of definition given.